

مفاهیم و روشهای آماری

۱

فنی و حرفه‌ای، گروه تحصیلی اداری مالی

(رشته حسابداری بازرگانی)

سال دوم دورهٔ آموزش متوسطه فنی و حرفه‌ای

۲۵۹/۴

بسم الله الرحمن الرحیم

مفاهیم و روشهای آماری(۱)

رشتهٔ حسابداری بازرگانی

گروه تحصیلی اداری مالی

زمینهٔ خدمات

شاخهٔ آموزش فنی و حرفه‌ای

شمارهٔ درس ۳۹۵۳

۰۰۱	مرآت، ابوالقاسم
/ ۴۲۲	مفاهیم و روشهای آماری (۱)/ مؤلف : ابوالقاسم مرآت. ــ تهران : شرکت چاپ و نشر کتابهای
م ۴۳۲ م	درسی ایران، ۱۳۹۴.
۱۳۹۴	۱۱۳ ص. ــ (آموزش فنی و حرفه‌ای؛ شمارهٔ درس ۳۹۵۳)
	متون درسی رشتهٔ حسابداری بازرگانی گروه تحصیلی اداری مالی، زمینهٔ خدمات.
	برنامه‌ریزی و نظارت، بررسی و تصویب محتوا : کمیسیون برنامه‌ریزی و تألیف کتابهای درسی رشتهٔ حسابداری بازرگانی دفتر تألیف کتابهای درسی فنی و حرفه‌ای و کاردانش وزارت آموزش و پرورش.
	۱. آمار. الف. ایران. وزارت آموزش و پرورش. دفتر تألیف کتابهای درسی فنی و حرفه‌ای و کاردانش. ب. عنوان. ج. فروست.

۱۳۹۴

همکاران محترم و دانش‌آموزان عزیز :

پیشنهادات و نظرات خود را دربارهٔ محتوای این کتاب به نشانی تهران ـ صندوق پستی شمارهٔ ۴۸۷۴/۱۵ دفتر تألیف کتابهای درسی فنی و حرفه‌ای و کاردانش، ارسال فرمایند.

پیام‌نگار (ایمیل) **info@tvoccd.sch.ir**

وب‌گاه (وب سایت) **www.tvoccd.sch.ir**

محتوای این کتاب براساس نظرهای هنرآموزان توسط مؤلف زیرنظر کمیسیون تخصصی برنامه‌ریزی و تألیف کتابهای درسی رشتهٔ حسابداری بازرگانی درسال ۸۷ ـ ۱۳۸۶ مورد بازبینی قرار گرفت و فصل ششم اضافه گردید.

وزارت آموزش و پرورش

سازمان پژوهش و برنامه‌ریزی آموزشی

برنامه‌ریزی محتوا و نظارت بر تألیف : **دفتر تألیف کتابهای درسی فنی و حرفه‌ای و کاردانش**

نام کتاب : **مفاهیم و روشهای آماری (۱) ـ ۳۵۹/۴**

مؤلف : **ابوالقاسم مرآت**

آماده‌سازی و نظارت بر چاپ و توزیع : **ادارهٔ کلّ نظارت بر نشر و توزیع مواد آموزشی**

تهران : خیابان ایرانشهر شمالی ـ ساختمان شمارهٔ ۴ آموزش و پرورش (شهید موسوی)

تلفن : ۹ـ ۸۸۸۳۱۱۶۱، دورنگار : ۸۸۳۰۹۲۶۶، کدپستی : ۱۵۸۴۷۴۷۳۵۹،

وب‌سایت : **www.chap.sch.ir**

رسام : **محمّد پریسای**

صفحه‌آرا : **صغری عابدی**

طراح جلد : **مریم کیوان**

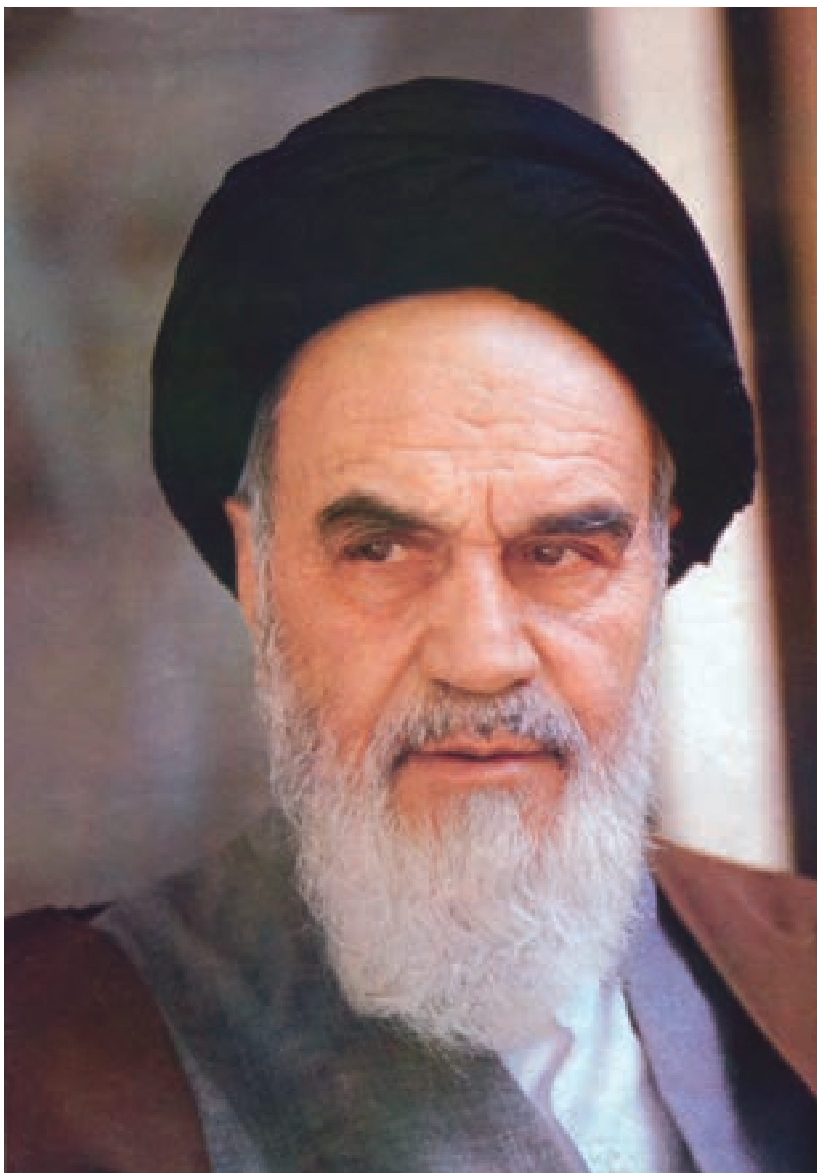
ناشر : **شرکت چاپ و نشر کتابهای درسی ایران : تهران ـ کیلومتر ۱۷ جادّهٔ مخصوص کرج ـ خیابان ۶۱ (داروپخش)**

تلفن : ۵ـ ۴۴۹۸۵۱۶۱، دورنگار : ۴۴۹۸۵۱۶۰، صندوق پستی : ۱۳۹ـ ۳۷۵۱۵

چاپخانه : **شرکت چاپ و نشر کتابهای درسی ایران «سهامی خاص»**

سال انتشار و نوبت چاپ : **چاپ شانزدهم ۱۳۹۴**

شابک ۹-۰۶۷۲-۰۵-۹۶۴ **ISBN 964-05-0672-9**



اول بایــد اخلاصتان را قــوی بکنید، ایمانتان را قــوی بکنید،... و این اخلاص و ایمان، شما را تقویت می‌کند و روحیهٔ شما را بالا می‌برد و نیروی شما جوری می‌شود که هیچ قدرتی نمی‌تواند (با شما) مقابله کند.

امام خمینی (ره)

فهرست

مقدمه

واژه آمار معادل لاتین Statistics است که از کلمه Status به معنای حالت، وضع یا موقعیت گرفته شده. کلمات Stato (دولت)، Statista(دولت‌شناس) نیز از همین خانواده هستند.

۱ــ مختصری دربارهٔ پیدایش، رشد و توسعهٔ آمار : علم آمار مانند علوم دیگر در نتیجهٔ نیازهای بشر به‌وجود آمده است و تاریخی طولانی دارد و از دورانهای گذشته تاکنون رشد و توسعه آن ادامه یافته است.

با پیدایش نخستین دولتها در تاریخ، آمار، هستی یافته است. در گذشته‌های دور هر دولت شکل‌گرفته، دارای دستگاههای اداره‌کننده و هم‌چنین نیروی ارتش بوده است. به همین جهت به آگاهی از تعداد جمعیّت، میزان دارایی و ثروت قلمرو حکومت خود نیاز داشته است تا امکانات و نیازمندیهای ضروری خود را فراهم سازد. دراین‌باره از زمانهای قدیم، سرشماریهای مختلف از جمعیت، تعداد دامها، محصولات کشاورزی و ... انجام می‌گرفته است که می‌توان از کشورهای مصر، چین، هند و بابل نام برد. در آغاز، عملیات آماری بسیار ساده بود و فقط بعضی از نمودهای اجتماعی را دربر می‌گرفت، رفته‌رفته با رشد تولید، ظهور روابط کالا و پول و درنتیجه با پیچیده‌تر شدن زندگی اقتصادی و امور دولتی، ثبت و نگهداری حساب نفوس، ثبت مشاغل، محدودهٔ اراضی، دامها و نظایر آن منظم‌تر و کاملتر شد، درنتیجه کارهای آماری نیز توسعه یافت. ظهور و توسعهٔ تجارت بین‌المللی باعث شد که نیاز به کسب اطلاعات دربارهٔ حکومتهای خارجی نیز بیشتر گردد، به‌طوری‌که سازمانهای دولتی و مؤسسات سرمایه‌داری مجبور به جمع‌آوری اطلاعات وسیع و مختلف دربارهٔ بازار کار و فروش کالا، منابع مواد خام و نظایر آنها شدند. این اطلاعات که بر پایهٔ نیازهای یادشدهٔ آن زمان به عنوان آمار رشد یافت، تنها یکی از پایه‌های آمار امروزی را تشکیل می‌داد ولی درطول زمان نیازها فزونی می‌گرفت و

آمار همچنان تکامل می‌یافت زیرا هنوز دانستنیها کافی نبودند. بدین‌جهت از قرن هفدهم به‌بعد گروهی از دانشمندان مکتبی به‌نامِ «حسابدانهای سیاسی» تشکیل دادند که هدف آن بیشتر کشف نظم روابط اجتماعی و مسایلی درارتباط با رشد و توسعهٔ سرمایه‌داری بود. هم‌زمان با این گروه، دانشمندان دیگری در آلمان مکتب آمار توصیفی یا دولت شناسی را تشکیل دادند، اینان سعی کردند که به‌طور همه‌جانبه با استفاده از اعداد، دولتها و کشورها را تفسیر نمایند. دراین سده‌ها رشد و توسعهٔ آمار، مراحل مختلفی را گذرانده، به یکی از تواناترین ابزارهای شناسایی روابط اجتماعی تبدیل شده بود. اواخر قرن نوزدهم و اوایل قرن بیستم دورهٔ رشد و تکامل سریع آمار است. در این زمان، روشهای آماری و استفاده از علوم ریاضی به‌طور وسیع توسعه یافته بود و در کنار آن به موازات آمار جمعیت، رشته‌های دیگر آمار اجتماعی و اقتصادی نیز رشد و توسعه می‌یافت.

۲ــ تعریف آمار به‌عنوان یک علم : تعریف موضوع آمار، مانند هر علم دیگر اهمیّت بسیار دارد. بحث دربارهٔ موضوع و روش آمار به‌طور متفاوت در کشورهای مختلف بیش از ۱۰۰ سال است که انجام می‌گیرد.

تعیین و تعریف موضوع یک علم، از تعیین کردن محتوای آن، جایگاه آن در بین علوم دیگر و چگونگی ارتباط آن با سایر علوم شکل می‌گیرد.

با توجّه به تاریخ توسعه و تکامل آمار، اکثر دانشمندان، آمار را به عنوان یک علم به‌صورت زیر تعریف نموده‌اند :

> «آمار یک علم مستقل است که جنبه‌های کمّی نمودهای اجتماعی را در ارتباط با کیفیت آنها، با هم مطالعه می‌کند.»

این تعریف، آمار را به عنوان یک علم مشخص کرده، اختلاف آن را از موضوعهای علوم دیگر بیان می‌کند. مطالعهٔ جنبه‌های کمّی نمودها در ارتباط با جنبه‌های کیفی آنها، نقشی اساسی در فرآیند تحلیل داده‌های آماری پیدا می‌کند. باید گفت که در روشهای مطالعهٔ نمودهای اجتماعی، تنها از مفاهیم و روشهای آماری استفاده نمی‌شود بلکه روشها و طریقه‌های دیگری نیز وجود دارد. از این‌رو، آمار کاربردی را به‌عنوان یک علم مستقل به‌صورت زیر تعریف می‌کنند :

> «آمار کاربردی علم مستقلی است که موضوع آن را تهیّه و تنظیم مفاهیم، روشها و الگوهای ریاضی موردنیاز تشکیل می‌دهد. همچنین طرح مسئله، جمع‌آوری داده‌های آماری، پردازش و استخراج آنها، به‌منظور تعبیر و استنتاج علمی و عملی نیز موضوع آمار کاربردی می‌باشد.»

روشهای آماری را می‌توان به دو بخش آمار توصیفی و آمار استنباطی تقسیم نمود :

آمار توصیفی، مجموعه‌ای از روشهاست که برای گردآوری، توصیف و بیان مقداری اطلاعات

مربوط به جامعه‌های مورد مطالعه، به‌کار می‌رود.

آمار استنباطی برای استنباط خصوصیّات یک جامعهٔ آماری از روی نتایج یک نمونه که از آن جامعه انتخاب می‌شود، به‌کار می‌رود. آمار استنباطی امکان می‌دهد تا جنبه‌های موردنظر جامعه‌های بزرگ آماری، از روی نمونه‌های نسبتاً کوچکی که نمایندهٔ آن جامعه‌ها هستند، توصیف شوند.

هدف کلی

آشنا کردن دانش‌آموزان با چگونگی توصیف و تعبیر مجموعهٔ داده‌های آماری (نتایج مشاهدات نسبت به یک خاصیّت مورد مطالعه در یک جامعه از اشیاء یا پدیده‌ها) با استفاده از مفاهیم و روشهای آماری.

فصل اول

مفاهیم اساسی

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند :

۱ــ مفاهیم اساسی آمار (جامعهٔ آماری، صفت مشخّصه، صفت متغیّر و انواع آن) را شرح دهید.

۲ــ مفاهیم اساسی آمار(جامعهٔ آماری،صفت مشخصه، صفت متغیّر و انواع آن) را با رشتهٔ فعالیت خود تطبیق دهد.

۳ــ مشاهدهٔ آماری را به عنوان یکی از روشهای تحقیق آماری توضیح دهد.

۴ــ یک صفت متغیّر را در یک جامعه مشاهده کند و نتایج آن را ارائه دهد.

۵ــ مراحل مطالعهٔ آماری را بیان کند.

۶ــ مفهوم اندازه‌گیری و انواع آن را شرح دهد و در مشاهدهٔ صفتهای متغیّر به‌کار برد.

مفاهیم اولیه

آمار نیز مانند علوم دیگر دارای مفاهیم اساسی است. این‌گونه مفاهیم را «مفاهیم اولیه» نیز می‌نامند که بدون تعریف‌اند و فقط توصیف می‌شوند.

جامعهٔ آماری

یکی از مفاهیم اولیه در آمار، مفهوم «جامعهٔ آماری» است که به‌صورت زیر بیان می‌شود :

«به همه افراد یا اشیایی که حداقل در یک خاصیّت باهم مشترک باشند «جامعهٔ آماری» گویند.»

به این‌ترتیب جامعهٔ آماری تنها به انسانها گفته نمی‌شود بلکه، جامعهٔ اشیا را نیز جامعه آماری می‌نامند.

«خاصیّت یا خواصی که بین اعضای جامعه مشترک هستند، «صفت مشخّصه» نامیده می‌شوند.»

مثلاً :

ــ جامعهٔ ایرانیان که صفت مشخّصهٔ آن «ایرانی بودن» است.

ــ جامعهٔ دانش‌آموزان ایرانی که صفات مشخّصه آن، «دانش‌آموز» و «ایرانی بودن» است.

ــ جامعهٔ روستاهای ایران که صفات مشخّصه آن «روستا بودن» و «متعلّق به ایران بودن» می‌باشد.

ــ جامعهٔ حسابداران که صفت مشخّصه آن «حسابدار بودن» است.

ــ جامعهٔ مهره‌های ساخته شده در یک دستگاه ماشینی که صفات مشخّصه آن یکی «مهره بودن» و دیگری «ساخته شدن به وسیلهٔ یک دستگاه ماشینی» است.

هریک از اعضای جامعه را «واحد جامعه» یا «فرد جامعه» می‌نامند.

مثلاً در جامعهٔ ایرانیان، یک فرد ایرانی، عضو یا واحد یا فرد آن جامعه می‌باشد یا در جامعهٔ دانش‌آموزان ایران، هر دانش‌آموز ایرانی یک واحد یا فردِ آن جامعه به‌حساب می‌آید.

تعداد واحدهای تشکیل‌دهندهٔ جامعه را، «حجم جامعه» گویند و آن‌را با حرف N نشان می‌دهند.

مثلاً اگر جمعیت ایران (درسال ۱۳۷۵) ۶۰ میلیون نفر باشد : $N = ۶۰,۰۰۰,۰۰۰$.

اگر تعداد مهره‌های ساخته شده در یک دستگاه ماشینی ۱۰۰,۰۰۰ عدد باشد : $N = ۱۰۰,۰۰۰$.

اعضای جامعهٔ آماری، می‌توانند محدود یا نامحدود باشند.

ــ **جامعهٔ محدود (متناهی)**: دارای حجم معیّن می‌باشد، به عبارت دیگر به تمامی اعضای آن می‌توان دسترسی داشت مانند جامعهٔ ایرانیان که در زمان معیّن (مثلاً سال ۱۳۷۵) دارای حجم $N = ۶۰,۰۰۰,۰۰۰$ است.

ــ **جامعهٔ نامحدود (نامتناهی)**: تعداد اعضای آن نامحدود است، به عبارت دیگر به تمام

اعضای آن نمی‌توان دسترسی داشت. مانند جامعه ستارگان، جامعه پرندگان و یا جامعهٔ ماهیان دریاها.

صفت متغیّر

عناصر جامعه بجز صفت مشخّصه که در همه اعضای آن مشترک است، خواص دیگری نیز دارند که باهم متفاوتند. این‌گونه خاصیتها یا صفات را که بین عناصر جامعه متفاوت هستند «صفات متغیّر» گویند. به عبارت دیگر،

صفتی که از یک عضو به عضو دیگر جامعه می‌تواند تغییر نماید، «صفت متغیّر» نامیده می‌شود.

مثلاً : در جامعهٔ ایرانیان، خصوصیاتی چون سن، طول قد، وزن، محل تولد، درآمد و ... صفات متغیّر می‌باشند. در ادامهٔ بحث، صفات متغیّر را مانند متغیّرها در ریاضی با حروف X، Y، Z و ... نشان خواهیم داد. صفات متغیّر بر دو نوعند : **صفت متغیّر کیفی، صفت متغیّر کمّی.**

ــ **صفت متغیّر کیفی:** صفتی است که آن را برای اعضای جامعه (در سطح علوم امروزی) با هیچ عدد یا کمیتی نتوان بیان کرد. مانند : جنس، نوع شغل، نوع بیماری، مرغوبیت کالا و ... .

ــ **صفت متغیّر کمّی:** صفتی است که آن را برای اعضای جامعه با عدد یا کمیت بتوان بیان کرد. مانند : طول قد افراد، درآمد افراد، تعداد اعضاء خانوار و ... .

از صفات کمّی : دو نوع مهم آن را مورد بحث قرار خواهیم داد :

ــ صفت کمّی گسسته (ناپیوسته)

ــ صفت کمّی پیوسته

صفت کمّی گسسته: صفتی است که دامنه تعریف آن مجموعه شمارا[1] (مجموعه نامحدود شمارش‌پذیر) یا هر زیر مجموعه‌ای از آن مجموعه می‌باشد.

مانند تعداد اعضای خانوار در جامعه خانوارهای یک منطقه، تعداد دانش‌آموزان یک دبیرستان، تعداد تصادفات اتومبیل در روز در یک شهر.

صفت کمّی پیوسته: صفتی است که دامنه تعریف آن را مجموعه اعداد حقیقی یا هر فاصله‌ای از آن مجموعه تشکیل می‌دهد.

مانند طول قد انسان، قطر داخلی مهره‌های ساخته شده توسط یک دستگاه که با یک عدد حقیقی بیان شوند.

۱ــ هر مجموعه نامحدود هم‌ارز با مجموعه اعداد طبیعی را مجموعه نامحدود شمارش‌پذیر یا مجموعه شمارا گویند.

مراحل مختلف انجام یک تحقیق آماری

تعیین هدف تحقیق: وقتی یک جامعهٔ آماری را مورد مطالعه قرار می‌دهیم، ابتدا باید هدف این بررسی را مشخص کنیم. پس از تعیین هدف، جامعهٔ آماری که مطالعه در آن انجام می‌گیرد باید کاملاً تعریف شود. همچنین صفت مشخّصه و صفات متغیّر مورد بررسی نیز باید دقیقاً روشن گردد.

برای مثال اگر بخواهیم درآمد خانوارها را در یک جامعه بررسی کنیم، قبل از هرچیز باید معلوم کنیم که هدف از مطالعه درآمد خانوارها چیست؟ ممکن است برای آگاهی از سطح درآمد خانوارها و مقایسه با هزینه‌های زندگی، این کار انجام پذیرد تا معلوم شود آیا بین درآمد و هزینهٔ خانوار در جامعهٔ مورد بررسی تعادلی وجود دارد یا خیر؟ بنابراین هدف از این بررسی، مقایسهٔ سطح درآمد خانوارها و هزینهٔ خانوارها به‌منظور آگاهی از وجود تعادل بین آنهاست. سپس جامعهٔ مورد بررسی را تعریف می‌کنیم. ممکن است جامعهٔ مورد بررسی، خانوارهای شهر اراک باشند. درنتیجه کلیه خانوارهایی که در شهر اراک ساکن هستند، جامعهٔ مورد بررسی را تشکیل می‌دهند. صفات متغیّر این بررسی می‌تواند میزان درآمد ماهانهٔ خانوارها، تعداد نفرات خانوارها، هزینهٔ ماهانه خانوارها و ... باشد.

پس از تعیین هدف و جامعهٔ مورد بررسی، برای انجام یک تحقیق آماری، سه مرحلهٔ اساسی زیر باید رعایت شود :

ــ **مشاهدهٔ آماری :** قبل از هر چیز باید صفت متغیر مورد مطالعه در یک جامعه برای اعضای آن جامعه مشخص گردد.

«مشخص‌کردن صفت متغیر برای هر یک از اعضای جامعه را مشاهده آماری می‌نامند».

نتایج هر مشاهده آماری اعضای جامعه در یک فرم آماری به‌نام پرسشنامه ثبت می‌گردد.

مثلاً با مراجعه به خانوارهای یک منطقه، ثبت اطلاعاتی نظیر درآمد، هزینه، تعداد اعضای خانوار و ... در واقع مشاهده صفت متغیر «درآمد خانوار»، «هزینه خانوار»، «تعداد اعضای خانوار» و ... می‌باشد. بسته به اینکه مشاهده آماری، تمامی اعضای جامعه را یا قسمتی از جامعه آماری را (زیرمجموعه‌ای از آن) دربرگیرد، دو نوع مشاهده آماری را از هم مجزا می‌کنند.

ــ مشاهده سرتاسری (تمام شماری)

ــ مشاهده غیرسرتاسری (یا مشاهده نمونه‌ای)

در مشاهده سرتاسری، تمامی اعضای جامعه، مورد مطالعه قرار می‌گیرند. لیکن به علل زیر امکان مشاهده سرتاسری در جامعه فراهم نمی‌باشد.

ــ هزینهٔ زیاد مشاهده

ــ دگرگونی اعضای جامعه

ــ نامحدود بودن حجم جامعه

مثلاً ،اگر بخواهیم تمامی اعضای جامعه انسانی را در یک کشور از نظر «سلامت بودن» مشاهده کنیم، آن‌چنان هزینه زیادی به بودجه کشور تحمیل می‌گردد، که هیچ کشوری تن به چنین مشاهده‌ای نمی‌دهد.

یا اگر بخواهیم طول عمر لامپهای تولید شده در یک کارخانه تولیدی را مطالعه کنیم، تمامی لامپهای ساخته شده باید به شبکه برق متصل گردد و پس از سوختن تمام لامپها، اطلاع راجع به طول عمر تمامی لامپها به‌دست خواهد آمد که چنین جامعه‌ای دیگر وجود ندارد و نابود شده است. چنین اطلاعاتی فاقد هرگونه ارزشی است یا مطالعه شدت حرارت ستارگان (درجه حرارت)، به‌علت نامحدود بودن حجم جامعه غیرممکن می‌باشد، در چنین شرایطی محقق ناچار به انتخاب جزئی از اعضای جامعه به عنوان نمونه می‌باشد که مشاهده غیرسرتاسری نامیده می‌شود.

ــ **پردازش داده‌های آماری:** اطلاعات به‌دست آمده در مرحلهٔ مشاهده (داده‌های آماری) را منظم کرده، آنها را برحسب مقادیر یا حالات صفت متغیّر گروه‌بندی می‌کنند و برای بیان اجمالی آنها از جداول و نمودارهای آماری کمک می‌گیرند. این عملیات را «پردازش داده‌های آماری» گویند.

ــ **مرحلهٔ تجزیه و تحلیل نتایج مشاهدات (داده‌های آماری):** پس از گروه‌بندی نتایج مشاهدات برحسب صفت متغیّر، مشخصه‌های عددی توصیف‌کننده صفت‌متغیّر در جامعه محاسبه می‌شود و بر پایهٔ آنها نسبت به چگونگی صفت متغیّر در جامعه، قضاوت و یا نتیجه‌گیری صورت می‌گیرد. این مرحله از مطالعه جامعهٔ آماری را «مرحلهٔ تجزیه و تحلیل آماری» می‌نامند.

اندازه‌گیری[1] و انواع اندازه‌گیری (مقیاسهای اندازه‌گیری)

گفته شد که برای مطالعه تغییرات صفت در جامعه، اکثر اوقات، نتیجهٔ مشاهده با اعدادی بیان می‌شوند که از اندازه‌گیری صفت متغیّر کمّی و یا از نسبت دادن اعداد به حالتهای صفت کیفی به‌دست می‌آیند. از این‌روی، روشن ساختن مفهوم اندازه‌گیری ضرورت پیدا می‌کند.

> نسبت دادن اعداد بر طبق قواعد معین به هر یک از اعضای جامعه را، اندازه‌گیری صفت متغیر در آن جامعه می‌نامند.

برای بیان تغییرپذیری صفت متغیر در جامعه، اکثر اوقات نتیجه مشاهده را با اعدادی بیان می‌کنند. هر یک از اعداد نسبت داده شده به اعضای جامعه را اندازهٔ آن صفت گویند.

۱ــ Measurement

در اینجا چهار نوع مهم اندازه‌گیری را (مقیاس اندازه‌گیری[1]) مورد بحث قرار می‌دهیم :

ــ مقیاس اسمی

ــ مقیاس ترتیبی

ــ مقیاس فاصله‌ای

ــ مقیاس نسبتی

ــ مقیاس اسمی[2]: در اندازه‌گیری به‌طریق «مقیاس اسمی» از اعداد به‌منظور جدا کردن عناصر طبقات مختلف، استفاده می‌شود. اعدادی که به نتایج مشاهدات نسبت داده می‌شوند، نقشِ «علامت» یا «نام» را برای طبقات دارند که مشاهدات در آنها قرار می‌گیرند. انتخاب نام «مقیاس‌اسمی» نیز از همین‌جا حاصل شده است. در این اندازه‌گیری، اعدادی که به عناصر نسبت داده می‌شوند، هیچ معنای کمّی ندارند و نمی‌توان بر روی آنها هیچ‌گونه عملیات جبری (چهارعمل اصلی) انجام داد.

فرض کنید می‌خواهیم افراد یک جامعه را برحسب جنس گروه‌بندی کنیم در این صورت مثلاً به هریک از مردها عدد ۱ و به هریک از زنها عدد ۲ را نسبت داده، آنها را به دو طبقه تقسیم می‌کنیم. نسبت دادن اعداد ۱ و ۲ به افراد جامعه، به‌منظور طبقه‌بندی آنها، اندازه‌گیری با مقیاس‌اسمی می‌باشد.

ــ مقیاس ترتیبی یا مقیاس رتبه‌ای[3]: در اندازه‌گیری با مقیاس ترتیبی، از اعداد برای مقایسهٔ عناصر ازنظر کوچکتر یا بزرگتر یا برابر بودن استفاده می‌شود که براساس نتایج این مقایسه‌ها نیز عناصر جامعه طبقه‌بندی می‌گردند. در این مقیاس، عناصر براساس اندازه‌های نسبی که برای آنها به‌دست می‌آید، مرتّب می‌شوند. بر روی اعداد مقیاس ترتیبی نیز هیچ‌گونه عملیات جبری انجام نمی‌گیرد.

مثال: در یک مؤسّسه تولیدی که ۱۵ نفر کارگر مشغول انجام یک نوع کار هستند، به‌منظور طبقه‌بندی آنها برحسب مهارت (صفت متغیّر)، می‌توان آنها را در چهار گروه : کارگر ساده، کارگر نیمه‌ماهر، کارگر ماهر و استادکار طبقه‌بندی نمود.

بدین‌منظور، به‌هریک از کارگران بسته به مهارت آنها، یکی از اعداد ۱، ۲، ۳ یا ۴ را نسبت می‌دهیم. نتایج به‌دست آمده به‌همان ترتیبی که مشاهده شده در زیر آورده شده است :

۳، ۲، ۳، ۳، ۱، ۴، ۱، ۳، ۲، ۲، ۳، ۳، ۱، ۴، ۲

براساس اعدادی که به آنها نسبت داده شده است، آنها را به طبقات زیر تقسیم‌بندی می‌کنیم :

(۴، ۴) و (۳، ۳، ۳، ۳، ۳، ۳) و (۲، ۲، ۲، ۲) و (۱، ۱، ۱، ۱)

۱ــ Scale of measurement

۲ــ Nominal scale

۳ــ Ordinal scale

نتیجه به‌دست آمده، طبقه‌بندی موردنظر است. هر طبقه، با رتبهٔ داده شده به اعضای آن، از طبقهٔ دیگر متمایز می‌گردد.

در بعضی موارد مقایسهٔ عناصر و یا مقایسهٔ چند گروه براساس امتیازهای داده شده به آنها با عدد برابر مشخص می‌گردند. مثلاً فرض کنیم به‌منظور رتبه‌بندی به‌صورت شاگرد اول، شاگرد دوم و شاگرد سوم در یک کلاس، پس از آزمونهای لازم، برای ۱۵ دانش‌آموز، نمرات زیر به‌دست آمده باشد :

۴، ۶، ۷، ۹، ۱۰، ۱۱، ۱۳، ۱۴، ۱۵، ۱۶، ۱۹، ۱۹، ۱۹، ۱۹، ۲۰

اگر در مشاهده، چند عضو اندازه‌های برابر داشته باشند، می‌گویند گروه همرتبه‌ها به‌وجود آمده است. بنابراین طبق تعریف فوق، بین دانش‌آموزان، یک گروه (۱۹، ۱۹، ۱۹، ۱۹) همرتبه وجود دارد.

در بعضی مواقع در تحلیل آماری صفت متغیّر در جامعه، نیاز است که عناصر در ترتیب معیّن با رتبه‌های یگانه مشخص گردند. در چنین مواردی، از رتبه‌های یکسان مُعدّل گرفته، آنها را ادغام می‌کنند و به‌جای رتبه هریک از آنها، معدّل رتبه‌ها را قرار می‌دهند. با یک مثال، این مطلب را روشن می‌سازیم.

در مثال بالا، عدد ۱ را به‌عنوان رتبه، به دانش‌آموزی که دارای بالاترین نمره است، نسبت می‌دهیم. عدد ۲ را به نمرهٔ پایین‌تر از رتبهٔ اول به‌عنوان رتبهٔ دوم نسبت می‌دهیم همچنین تا آخر. به دانش‌آموز با نمرهٔ ۲۰ رتبه اول را می‌دهیم ولی به دانش‌آموز با نمرهٔ ۱۹ رتبهٔ دوم را می‌دهیم. در اینجا ملاحظه می‌شود که ۴ دانش‌آموز، نمرهٔ ۱۹ گرفته‌اند، بنابراین باید به آنها رتبهٔ دوم تا رتبهٔ پنجم را اختصاص دهیم. اما از آنجا که ارزش نمرهٔ آنها با هم یکسان است از رتبه ۲ تا ۵ معدّل گرفته و عدد به‌دست آمده را برای نفرات دوم تا پنجم قرار می‌دهیم.

$$\frac{۲+۳+۴+۵}{۴}=۳/۵$$

درواقع برای نفرات دوم تا پنجم رتبهٔ ۳/۵ را درنظر می‌گیریم. درمورد سایر نمرات که تکرار شده‌اند نیز به‌همین ترتیب عمل می‌کنیم. درنتیجه دانش‌آموزان موردنظر، برحسب رتبه به‌صورت زیر قرار خواهند گرفت :

۱، ۳/۵، ۳/۵، ۳/۵، ۳/۵، ۶، ۷، ۸، ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵

ــ مقیاس فاصله‌ای[1]: در اندازه‌گیری به شیوهٔ «مقیاس فاصله‌ای» عناصر مورد اندازه‌گیری نه‌تنها ازنظر ترتیب اندازه می‌توانند مقایسه شوند، بلکه برحسب طول فاصله‌ها بین اندازه‌های عناصر نیز می‌توانند مقایسه و طبقه‌بندی گردند. «مقیاس فاصله‌ای» از مفهوم «واحد فاصله» استفاده می‌کند و

۱ــ Interval scale

بدین‌جهت می‌تواند فاصله را بین دو اندازه برحسب تعداد این «واحد فاصله»ها بیان کند.

مثال روشن برای اندازه‌گیری به‌طریق مقیاس فاصله‌ای، مقیاسی است که با آن شدّت حرارت را نشان می‌دهیم. «واحد فاصله» در اینجا برای اندازه‌گیری «درجه» است. مقدار عددی شدّت حرارت به‌طور ساده، نتیجهٔ مقایسه با یک نقطهٔ دلخواه است که آن‌را «درجهٔ صفر» می‌نامند. در اندازه‌گیری با مقیاس فاصله‌ای، به وجود «**نقطهٔ صفر**» و «**واحد فاصله**» نیاز می‌باشد. ولی اهمیّت ندارد که این نقطه با چه نوع اندازه‌گیری مشخص شده است و یا چه فاصله‌ای به‌عنوان «واحد فاصله» تعریف شده است. شدّت حرارت گاهی با مقیاس سانتیگراد و گاهی با مقیاس فارنهایت اندازه‌گیری می‌شود که دارای نقطهٔ صفر و واحد فاصله متفاوت هستند.

برخلاف مقیاس اسمی و مقیاس ترتیبی که روی نتایج اندازه‌گیری، عملیات جبری (چهارعمل اصلی) غیرممکن است، در مقیاس فاصله‌ای انجام بعضی از عملیات جبری روی اعداد به‌دست آمده از اندازه‌گیری، امکان‌پذیر است. مثلاً در این مقیاس می‌توان گفت که ۵ − ۹ = ۱۴− ۱۸ و برای هر دوطرف تساوی می‌توان گفت، اندازه اولی چهار واحد بیشتر از دومی است. ولی در این مقیاس نمی‌توان گفت که اندازهٔ یک عنصر چند برابر اندازهٔ عنصر دیگر است. مثلاً در اندازه‌گیری شدّت حرارت با مقیاس سانتیگراد نمی‌توان گفت که درجهٔ حرارت ۵۰ دو برابر درجهٔ حرارت ۲۵ می‌باشد. مثال دیگر مقیاس تقویمی است که در سال شمسی مبدأ آن، هجرت حضرت محمد(ص) از مکّه به مدینه می‌باشد و در سال میلادی مبدأ آن، میلاد حضرت مسیح (ع) است.

ــ **مقیاس نسبتی**[1]: این مقیاس هم زمانی که اعداد به‌دست‌آمده از اندازه‌گیری، ازنظر ترتیب و فاصله بین دو مقدار، دارای اهمیت هستند، به‌کار می‌رود و هم وقتی که نسبت دو عدد به‌دست آمده از اندازه‌گیریها، دارای اهمیت باشد. به عبارت دیگر، مقیاس نسبتی امکان می‌دهد که تعیین کنیم یک اندازه از اندازهٔ دیگر چقدر بیشتر یا کمتر است و نیز می‌توان تعیین کرد این اندازه برای یک عنصر چند برابر اندازهٔ عنصر دیگر است. وقتی اندازه‌گیری به‌منظور مقایسهٔ دو عنصر، به‌صورت نسبت اندازه‌های آنها باشد، استفاده از اندازه‌گیری با مقیاس نسبتی اجتناب‌ناپذیر است. تنها اختلاف اساسی بین مقیاس نسبتی و مقیاس فاصله‌ای در این است که برای مقیاس نسبتی اندازه طبیعی به نام «صفر مطلق» وجود دارد، در صورتی که در اندازه‌گیری بر طبق مقیاس‌فاصله‌ای، اندازهٔ صفر به‌طور دلخواه تعیین می‌گردد. در مقیاس نسبتی نیز مانند مقیاس‌فاصله‌ای، «واحد فاصله» بین دو اندازه برای عناصر به‌طور دلخواه تعیین می‌شود. برخلاف مقیاس‌فاصله‌ای که انجام عمل تقسیم روی اعداد

۱ــ Ratio scale

به‌دست‌آمده از اندازه‌گیری غیرممکن می‌باشد، در مقیاس نسبتی انجام این عمل امکان‌پذیر است و می‌توان بین نسبتهایی از اعداد که از اندازه‌گیریها به‌دست می‌آیند، تساوی برقرار کرد. مثلاً در این مقیاس می‌توان گفت : $\frac{۱}{۲} . \frac{۶}{۱۲}$ که نتیجهٔ وجود نقطهٔ صفر تثبیت شده در این مقیاس می‌باشد و برای دو اندازه ۱۲ و ۶ می‌توان گفت که اندازه اولی ۲ برابر اندازه دومی است.

صفتهای متغیّر کمّی در مقیاس فاصله‌ای و نسبتی اندازه‌گیری می‌شوند ولی صفتهای متغیّر کیفی در سطح مقیاس اسمی و ترتیبی اندازه‌گیری می‌شوند.

علامت زیگما (Σ)[1] و کاربردِ آن در محاسبات

از آنجا که در بسیاری از فرمولهای آماری از علامت زیگما (.) استفاده می‌شود، به‌منظور کاربرد صحیح آن، ضروری است با خصوصیّات این علامت آشنا شویم.

قاعدهٔ ۱: برای نشان دادن عمل جمع چند عدد، از حرف یونانی . (زیگما) استفاده می‌شود. اگر n مقدار $x_۱$ ، $x_۲$ ، $x_۳$ ، .... ، x_n در اختیار باشد، حاصل جمع آنها با استفاده از حرف زیگما می‌تواند به‌صورت $\overset{n}{\underset{i.۱}{}} x_i$ (بخوانید زیگمای x_i از ۱ تا n) نشان داده می‌شود. بنابراین داریم :

$$\overset{n}{\underset{i.۱}{}} x_i \;.\; x_۱ + x_۲ + x_۳ + \quad + x_n$$

وقتی دامنه مقادیر از محتوای مسأله آشکار باشد می‌توان بدون نوشتن اعداد برای اندیس i حاصل جمع n عدد را به‌طور ساده‌تر به‌صورت x_i نوشت.

قاعدهٔ ۲: ضرب مقادیر x_i در عدد ثابت a

اگر همه اندازه‌های صفت را در عدد ثابت a ضرب کرده، سپس آنها را با هم جمع کنیم، نتیجه را با استفاده از علامت . می‌توان به‌صورت زیر ساده کرد :

$$ax_i \;.\; ax_۱ + ax_۲ + ax_۳ + \quad + ax_n \;.\; a(x_۱ + x_۲ + \quad + x_n) \;.\; a \quad x_i$$

مثال: می‌خواهیم مقادیر متغیّر x را که ۱۴، ۱۲، ۱۰، ۸، ۶، ۴، ۲ می‌باشند، در عدد ۵ . a ضرب کرده، حاصل ضربها را جمع کنیم.

با استفاده از قاعدهٔ ۲ خواهیم داشت :

$$۵x_i \;.\; ۵ . ۲ + ۵ . ۴ + ۵ . ۶ + ۵ . ۸ + ۵ . ۱۰ + ۵ . ۱۲ + ۵ . ۱۴$$

$$. \; ۵(۲ + ۴ + ۶ + ۸ + ۱۰ + ۱۲ + ۱۴) \;.\; ۵ . ۵۶ \;.\; ۲۸۰$$

۱ــ حرف زیگما، معادل حرف S انگلیسی در زبان یونانی می‌باشد که حرف اول واژه Sum به معنای مجموع است.

قاعدهٔ ۳: حاصل‌جمع مجموع چند متغیّر برابر است با مجموع حاصل‌جمع آن چند متغیّر :

$$\sum_{i=1}^{n}(x_i+y_i)=\sum x_i+\sum y_i$$

$$\sum_{i=1}^{n}(x_i+y_i)=(x_1+y_1)+(x_2+y_2)+\cdots+(x_n+y_n)$$

$$=(x_1+x_2+\cdots+x_n)+(y_1+y_2+\cdots+y_n)$$

$$=\sum x_i+\sum y_i$$

قاعدهٔ ۴: اگر a مقدار ثابتی باشد، خواهیم داشت :

$$\sum_{i=1}^{n}a=na$$

$$\sum_{i=1}^{n}a=a+a+\cdots+a=na$$

سؤالها و تمرینها

۱ــ آمار را به‌عنوان یک علم تعریف کنید.

۲ــ جامعهٔ آماری چیست؟

۳ــ صفت مشخّصه جامعهٔ آماری چیست؟

۴ــ فرد جامعه چیست؟

۵ــ حجم جامعهٔ آماری چیست؟

۶ــ جامعهٔ محدود و جامعهٔ نامحدود را تعریف کرده، برای هرکدام مثالی بیاورید.

۷ــ صفت متغیّر چیست؟

۸ــ صفت متغیّر کمّی چیست؟ تعریف کنید و یک مثال بیاورید.

۹ــ صفت متغیّر کیفی چیست؟ تعریف کنید و یک مثال بیاورید.

۱۰ــ انواع صفت کمّی را نام برده، برای هریک مثالی بیاورید.

۱۱ــ برای به‌دست آوردن داده‌های آماری، چه اقدامی باید انجام گیرد؟

۱۲ــ مشاهدهٔ آماری چیست و به چند طریق انجام می‌گیرد؟

۱۳ــ مشاهدهٔ سرتاسری چه نوع مشاهده‌ای است؟ مشکلات آن را نام ببرید.

۱۴ــ مشاهدهٔ غیر سرتاسری چه نوع مشاهده‌ای است و در چه مواقعی الزاماً به‌کار می‌رود؟

۱۵ــ در انجام یک تحقیق آماری چه مراحلی را باید رعایت نمود؟

۱۶ــ پردازش آماری شامل چه اقداماتی است؟

۱۷ــ تحلیل آماری به چه منظور انجام می‌گیرد؟

۱۸ــ اندازه‌گیری صفت متغیر چیست؟

۱۹ــ مقیاس اندازه‌گیری چیست؟

۲۰ــ چند نوع مقیاس اندازه‌گیری مهم مورد استفاده قرار می‌گیرد؟

۲۱ــ در اندازه‌گیری به شیوهٔ «مقیاس‌اسمی» اعداد نسبت داده شده به عناصر، چه معنایی دارند؟

۲۲ــ آیا روی اعداد نسبت داده شده در مقیاس اسمی، می‌توان عملیات جبری را انجام داد؟ چرا؟

۲۳ــ در مقیاس ترتیبی، چگونه اعداد به عناصر مورد اندازه‌گیری نسبت داده می‌شوند؟ آیا روی این اعداد، می‌توان عملیات جبری را انجام داد؟ چرا؟

۲۴ــ تفاضل دو اندازهٔ ۷ و ۴ را در اندازه‌گیری به‌طریق مقیاس فاصله‌ای چگونه تعبیر می‌کنید؟

۲۵ــ اندازه‌گیری به شیوهٔ مقیاس نسبتی، چه خصوصیّاتی از اعداد به‌دست آمده در این نوع اندازه‌گیری را درنظر می‌گیرد؟ درچه مواردی استفاده از مقیاس نسبتی در اندازه‌گیریها، اجتناب‌ناپذیر است؟

۲۶ــ اختلاف اساسی بین اندازه‌گیری به شیوهٔ مقیاس‌نسبتی و مقیاس‌فاصله‌ای در چیست؟

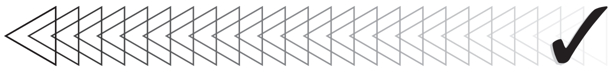
۲۷ــ نماد (!)، به چه منظور در آمار وارد شده است؟

۲۸ــ چه قواعدی در محاسبات برای علامت ! وجود دارد؟

۲۹ــ اگر x_1 ، x_2 ، x_3 ، x_4 ، x_5 به‌ترتیب اعداد زوج متوالی ۲، ۴، ۶، ۸، ۱۰ باشند، مقادیر $\sum_{i=1}^{n} x_i$ ، $\sum_{i=2}^{4} x_i$ ، $\sum x_i^2$ و $(\sum x_i)^2$ را محاسبه نمایید.

۳۰ــ در تمرین ۲۹ اگر $a=10$ باشد، مقادیر : $\sum_{i=1}^{n} ax_i$ و $\sum_{i=1}^{n} a^2x_i^2$ و $\sum_{i=1}^{n} (x_i+a)$ و $\sum_{i=1}^{n} (x_i-a)^2$ را محاسبه نمایید.

خودآزمونهای چهارگزینه‌ای



۱ــ کدام‌یک از گزینه‌های زیر معرّف جامعهٔ آماری است؟

الف ــ اعضای آن دارای خواص مختلفی باشند.

ب ــ همهٔ اعضای آن حداقل یک خاصیت مشترک داشته باشند.

ج ــ اعضای آن دارای چند صفت متغیّر باشند.

د ــ اعضای آن متعلّق به یک گروه همگن باشند.

۲ــ کدام تعریف برای صفت مشخّصه جامعه صحیح است؟

الف ــ صفتی است که از فردی به فرد دیگر تغییر می‌کند.

ب ــ صفتی است که برای اعضای با عدد یا کمیت قابل بیان باشد.

ج ــ صفت مشترک برای افراد جامعه است.

د ــ صفتی است که قابل شمارش است.

۳ــ کدام‌یک از گزینه‌های زیر تعریف صفت متغیّر کمّی است؟

الف ــ صفتی است که با مجموعه اعداد گویا بیان شود.

ب ــ صفتی است که با عدد یا کمیت قابل بیان باشد.

ج ــ با عدد یا کمیت نتوان آن را بیان کرد.

د ــ با اعداد مثبت بیان شود.

۴ــ کدام‌یک از گزینه‌های زیر تعریف صفت متغیر کیفی است؟

الف ــ با مجموعه اعداد صحیح بیان شود.

ب ــ با هیچ عدد یا کمیتی بیان نشود.

ج ــ با مجموعه اعداد گویا بیان نشود.

د ــ با عدد یا کمیت بیان شود.

۵ــ کدام‌یک از گزینه‌های زیر، تعریف صفت کمّی پیوسته است؟

الف ــ صفتی است که دامنه تعریف آن مجموعه اعداد شمارش‌پذیر باشد.

ب ــ صفتی است که دامنه تعریف آن مجموعه اعداد حقیقی یا هر فاصله‌ای از آن باشد.

ج ــ صفتی است که دامنه تعریف آن مجموعه اعداد گویا باشد.

د ــ صفتی است که با مجموعه اعداد اصم بیان شود.

۶ــ مشاهدهٔ آماری یعنی :

الف ــ مشخص‌کردن صفت متغیر برای هر یک از اعضای جامعه

ب ــ ثبت حالات صفات کیفی در جامعه

ج ــ ثبت مقدار صفت متغیّر کمّی برای اعضای جامعه

د ــ اندازه‌گیری صفت متغیّر کمّی برای اعضای جامعه

۷ــ مهمترین مشکلات مشاهدهٔ سرتاسری کدام‌اند؟

الف ــ عدم دسترسی به تمامی اعضای جامعه

ب ــ نداشتن پرسنل مجرب برای آمارگیری

ج ــ هزینه زیاد، نامحدود بودن جامعه و بعضاً انهدام جامعه

د ــ طولانی شدن زمان پردازش داده‌های آماری

۸ــ کدام‌یک از گزینه‌های زیر، مفهوم اندازه‌گیری برای صفت متغیّر جامعه را بیان می‌کند.

الف ــ وسیله‌ای که با آن صفات را اندازه‌گیری می‌کنند.

ب ــ نسبت دادن اعداد برطبق قواعد معیّن به اعضای جامعه برای مشخص شدن صفت متغیر آن

ج ــ اندازه‌گیری طول صفات متغیّر در جامعه

د ــ نسبت اندازه صفات متغیّر در جامعه

۹ــ روی کدام‌یک از مقیاسهای اندازه‌گیری، می‌توان عملیات جبری (چهارعمل اصلی) را انجام داد؟

الف ــ مقیاس اسمی　　　ب ــ مقیاس فاصله‌ای

ج ــ مقیاس رتبه‌ای　　　د ــ مقیاس نسبتی

۱۰ــ ! در ضرب اعداد چه مفهومی دارد؟

الف ــ فاکتورگیری از ضرایب ؛ Xها (متغیّرها)

ب ــ حاصل جمع، حاصلضربهای عدد ثابت در متغیّرها

ج ــ تفکیک ضرایب از مقادیر صفت

د ــ تفکیک مقادیر صفت از ضرایب

فصل دوم

گروه‌بندی و پردازش داده‌های آماری

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند :

۱ــ توضیح دهد که پردازش آماری به چه منظوری انجام می‌گیرد.

۲ــ مفهوم گروه‌بندی و نقش آن‌را برای توصیف آماری توضیح دهد.

۳ــ نتایج مشاهدات را در یک جامعهٔ آماری گروه‌بندی نماید.

۴ــ مفهوم فراوانی را توضیح دهد.

۵ــ توزیع فراوانیهای صفت متغیّر را در جامعه بیان کند.

۶ــ جدول «توزیع فراوانیهای صفت متغیّر» را با استفاده از نتایج مشاهدات، تشکیل دهد.

۷ــ نمودار جدول «توزیع فراوانیهای صفت متغیّر» را رسم کند.

۸ــ مفهوم «فراوانی انباشته» را تعریف نماید.

۹ــ براساس نتایج مشاهدات، فراوانیهای انباشته را محاسبه کرده، توزیع فراوانیهای انباشته را تشکیل دهد و نقش آن‌را به‌عنوان مشخّصهٔ صفت متغیّر توضیح دهد.

۱۰ــ لزوم واردکردن فاصله‌ها را برای مقادیر یک صفت متغیّر نشان دهد.

۱۱ــ لزوم واردکردن «مفهوم چگالی فراوانیها» را بیان کند.

۱۲ــ جدول توزیع چگالی فراوانی را تشکیل داده، نمودار آن را نیز رسم کند.

پردازش داده‌های آماری ــ توزیع صفت متغیّر

پردازش داده‌های آماری[1]

داده‌های آماری که در مرحلهٔ مشاهده به‌دست آمده است، اعداد یا ارقامی هستند که هیچ‌گونه

۱ــ Data processing

مفهوم خاصی نداشته و نیاز به پردازش دارند تا آنها را به‌صورت یک مجموعهٔ ادغام شده درآوَردْ. پردازش نتایج مشاهدات، شامل منظم کردن، طبقه‌بندی یا گروه‌بندی داده‌ها، تشکیل جداول، محاسبه سرجمعها و مشخصه‌های عددی است که آن‌را «مرحلهٔ پردازش» می‌نامند. بنابراین هدف از پردازش، ادغام نتایج مشاهدات، یا متراکم کردن اطلاعات، به‌دست‌آوردن مشخصه‌های کلی به‌منظور توصیف اجمالی خصوصیات جامعه‌ها می‌باشد.

توزیع صفت متغیّر

مطالعهٔ صفت درجامعه، بیشتر به‌منظور آشکارساختن تغییرپذیری صفت متغیّر صورت می‌گیرد. برای این‌کار لازم است عناصر جامعه در گروههای یکسان دسته‌بندی شوند. گروه‌بندی نتایج مشاهدات، یکی از اساسی‌ترین روشهای پردازش و تحلیل اولیه اطلاعات آماری است.

اگر صفت متغیر مورد بررسی کمی گسسته باشد، مقادیر مشاهده شده یکسان را که با یک عدد بیان می‌شوند به‌عنوان یک گروه درنظر گرفته، تعداد آنها را مشخص می‌کنیم. سپس تمامی مقادیر مختلف صفت متغیر را به‌ترتیب صعودی مرتب نموده در مقابل هریک از این مقادیر تعداد آنها (فراوانی[1]) را می‌نویسیم. نتیجه گروه‌بندی به‌صورت زیر به‌دست می‌آید:

$$X: \; x_1 \;\; x_2 \;\; \ldots \;\; x_k$$

$$F_i: \; F_1 \;\; F_2 \;\; \ldots \;\; F_k$$

دو مجموعه مقادیر صفت متغیر X و فراوانیها را با هم «توزیع فراوانیهای صفت متغیر» یا به‌طور ساده «توزیع صفت متغیر» می‌نامند.

تعداد عناصر را در هر گروه، «فراوانی مطلق[2]» آن گروه می‌نامند و برطبق معمول آن‌را با F_i نشان می‌دهند.

واضح است که برای فراوانیهای مطلق در یک توزیع صفت، همواره تساوی زیر برقرار است:

$$F_1 + F_2 + F_3 + \ldots + F_k = N \qquad (1)$$

به‌عبارت دیگر، مجموع فراوانیهای مطلق صفت متغیّر در جامعه، همواره برابر N یعنی حجم جامعه می‌باشد.

1ـ Frequency

2ـ Absolute Frequency

توزیع صفت متغیّر، چگونگی تغییرات فراوانیها را برحسب تغییرپذیری صفت متغیّر منعکس می‌نماید. توزیع صفت متغیّر را با جدول نیز به‌صورت زیر می‌توان بیان نمود :

جدول ۱

X	$x_۱$	$x_۲$	$x_۳$	x_k	–
$\mathbf{F_i}$	$F_۱$	$F_۲$	$F_۳$	F_k	N

چنین جدولی را «جدول توزیع فراوانیهای مطلق» می‌نامند.

مثال ۱: فرض کنیم جامعهٔ مورد مطالعه N=۲۰ خانوار باشد که صفت متغیّر X در آن تعداد اعضای هر خانوار است. برای هر خانوار، تعداد اعضای آن با پرسش تعیین شده و نتایج مشاهدات به‌صورت زیر به‌دست آمده است :

X:۲,۴,۳,۴,۴,۵,۲,۲,۴,۵,۵,۳,۴,۴,۳,۳,۲,۴,۳,۵

اگر مقادیر مختلف صفت متغیّر X را که به‌ترتیب ۲ ، ۳ ، ۴ و ۵ می‌باشند،در ۴ گروه طبقه‌بندی نموده، تعداد تکرار آنها (فراوانیها) را در برابر آنها قرار دهیم،گروههایی به‌صورت زیر به‌دست می‌آیند (جدول ۲) :

جدول ۲

X تعداد اعضای خانوار	۲	۳	۴	۵	–
$\mathbf{F_i}$ **تعداد خانوار(فراوانی)**	۴	۵	۷	۴	. F_i = ۲۰

اگر فراوانیهای مطلق (F_i) را برحجم جامعه (N) تقسیم کنیم، توزیع فراوانیهای نسبی[۱] را به‌دست خواهیم آورد. فراوانیهای نسبی را با نماد f_i نشان خواهیم داد که از رابطهٔ زیر به‌دست می‌آید :

$$f_i = \frac{F_i}{N} \qquad (۲)$$

فراوانی نسبی گروه iام در جدول توزیع فراوانیها، سهم هریک از مقادیر صفت متغیّر را در کل جامعه نشان می‌دهد.

۱ــ Relative Frequency

حاصل جمع فراوانیهای نسبی در جدول توزیع فراوانی، همواره برابر ۱ می‌باشد.

$$f_1 + f_2 + f_3 + \cdots + f_k = \sum_{i=1}^{k} f_i = 1 \qquad (3)$$

توزیع فراوانیهای نسبی صفت متغیّر را با جدول نیز می‌توان نشان داد :

جدول ۳ــ توزیع فراوانیهای نسبی صفت متغیّر X

X	x_1	x_2	x_3	$\cdots$	x_k	–
f_i	f_1	f_2	f_3	$\cdots$	f_k	$\sum f_i = 1$

برای مثال ۱، توزیع فراوانیهای نسبی را محاسبه می‌کنیم (جدول ۴) :

جدول ۴

X تعداد اعضای خانوار	۲	۳	۴	۵	–
f_i نسبت خانوارها در جامعه	$\frac{4}{20}$	$\frac{5}{20}$	$\frac{7}{20}$	$\frac{4}{20}$	۱

مثال ۲: سن کارگران یک کارخانه نساجی مورد پرسش واقع شده، نتایج به‌دست آمده در جدول زیر، گروه‌بندی شده است. می‌خواهیم جدول توزیع فراوانی نسبی سن کارگران را به‌دست آوریم :

جدول ۵

X سن	۱۹	۲۰	۲۲	۲۵	۲۶	۳۱	۳۴	۳۷	–
F_i فراوانی مطلق	۷	۶	۹	۱۰	۱۵	۱۲	۸	۱۳	۸۰

کافی است هریک از فراوانیهای مطلق (F_i) را به N (در اینجا $N = 80$) تقسیم نماییم تا فراوانی نسبی هر گروه از سن کارگران به‌دست آید. نتیجهٔ محاسبات در جدول ۶ آورده شده است :

جدول ۶

X سن	۱۹	۲۰	۲۲	۲۵	۲۶	۳۱	۳۴	۳۷	–
f_i فراوانی نسبی	$\frac{7}{80}$	$\frac{6}{80}$	$\frac{9}{80}$	$\frac{10}{80}$	$\frac{15}{80}$	$\frac{12}{80}$	$\frac{8}{80}$	$\frac{13}{80}$	$\frac{80}{80} = 1$

ملاحظه می‌شود جمع فراوانیهای نسبی برابر ۱ خواهد شد.

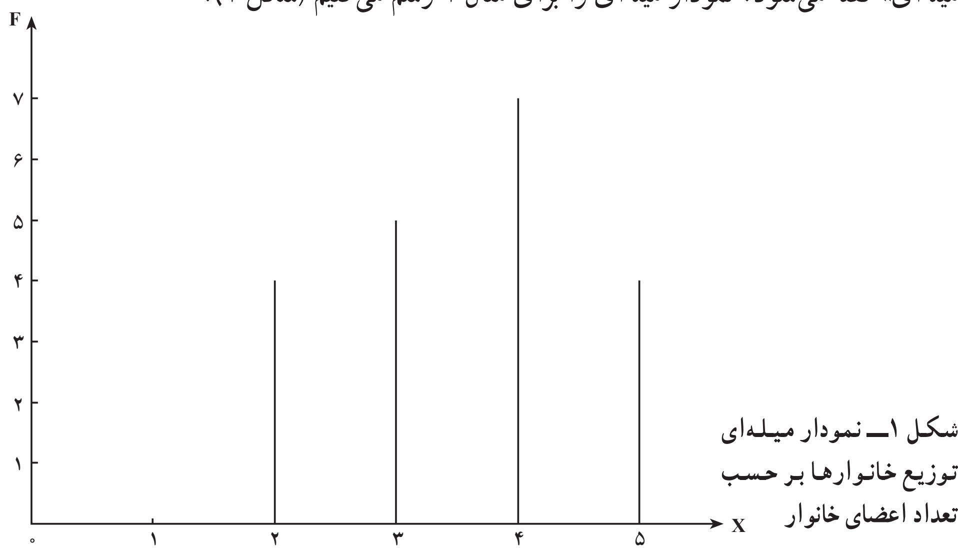
مفهوم فراوانی نسبی بهتر از فراوانی مطلق می‌تواند وضعیّت صفت را در جامعه مشخص سازد، برای مثال اگر بگوییم $\frac{۱۰}{۸۰}$ کارگران سنی برابر ۲۵ سال دارند، بهتر است از اینکه بگوییم در این کارخانه ۱۰ کارگر ۲۵ ساله وجود دارد. چون فراوانی مطلق معلوم نمی‌کند از چند کارگر موجود در این کارخانه، ۱۰ نفر ۲۵ سال سن دارند، ولی فراوانی نسبی دقیقاً آن‌را مشخص می‌کند.

بیان هندسی توزیع صفت متغیّر (نمودار میله‌ای[۱] ــ نمودار چندگوش[۲])

نمایش تغییرپذیری صفت متغیّر با جداول توزیع فراوانیها و پی‌بردن به خصوصیات آن به‌ویژه وقتی این‌گونه جداول مفصل باشند، ساده نیست به‌طوری‌که از روی آنها به‌آسانی نمی‌توان تغییرات فراوانیها را از یک مقدار به مقدار دیگر صفت فهمید.

برای آشکار کردن آنها در آمار، غالباً از نمودار میله‌ای و نمودار چندگوش، استفاده می‌شود. (البته این‌کار بیشتر اوقات زمانی که صفت متغیّر کمی گسسته باشد صورت می‌گیرد).

برای ساختن نمودار میله‌ای در دستگاه مختصات (قائم)، بر روی محور طولها، مقادیر متغیّر X و بر روی محور عرضها، فراوانیها، قرار داده می‌شوند. هر زوج X_i و F_i از توزیع فراوانیهای صفت متغیّر، در دستگاه مختصات به‌صورت یک نقطه با طول X_i و عرض F_i نشان داده می‌شود. پس از رسم تمامی نقاط بر روی دستگاه مختصات، از نقاط به‌دست آمده، عمودهایی به محور Xها رسم می‌کنیم. به‌این‌ترتیب تصویری به‌دست می‌آید که از چند میله تشکیل شده است و به آن «نمودار میله‌ای» گفته می‌شود. نمودار میله‌ای را برای مثال ۱ رسم می‌کنیم (شکل ۱).



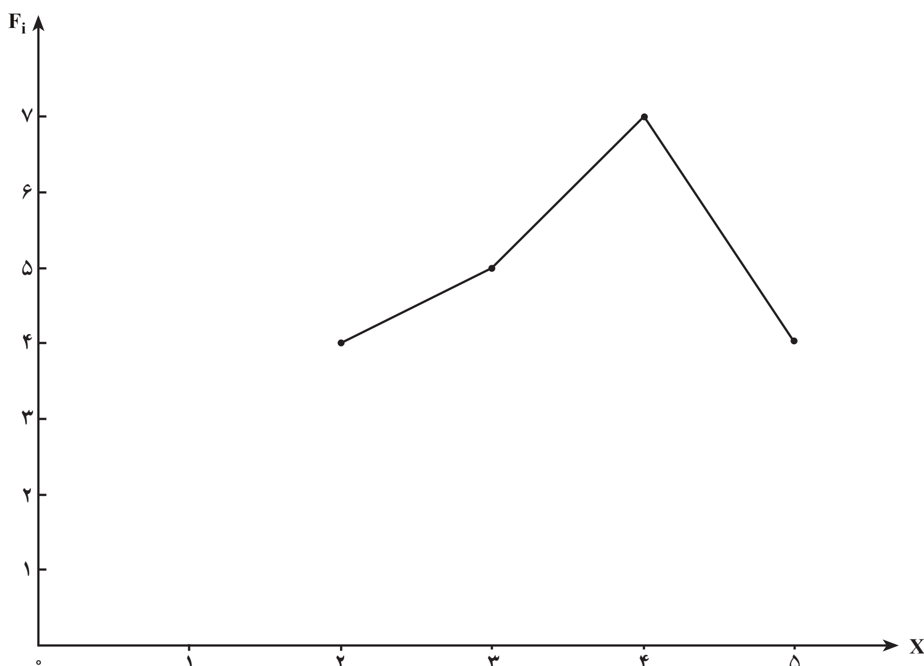
شکل ۱ــ نمودار میله‌ای توزیع خانوارها بر حسب تعداد اعضای خانوار

۱ــ Bar Diagram

۲ــ Polygon

با نمودار میله‌ای به سهولت می‌توان دریافت که خانوارهای ۴نفره تعدادشان بیش‌از همه است و بعد از آن خانوارهای ۳ نفره و سپس خانوارهای ۲ و ۵ نفره به تعداد یکسان در این جامعه وجود دارند.

برای رسم نمودار چندگوش، کافی است نقاط به‌دست آمده در دستگاه مختصات را با خطوط مستقیم به‌ترتیب به‌هم وصل نمود. درنتیجه شکلی حاصل می‌شود که از چند گوشه (زاویه) تشکیل شده است و به‌همین دلیل به آن «نمودار چندگوش توزیع فراوانیها» گفته می‌شود. این نوع نمودار بیشتر برای مقایسهٔ توزیع فراوانیها در دو یا چند جامعه باهم مورد استفاده قرار می‌گیرد. چندگوش توزیع فراوانیها را برای مثال ۱ رسم می‌کنیم (شکل ۲) :



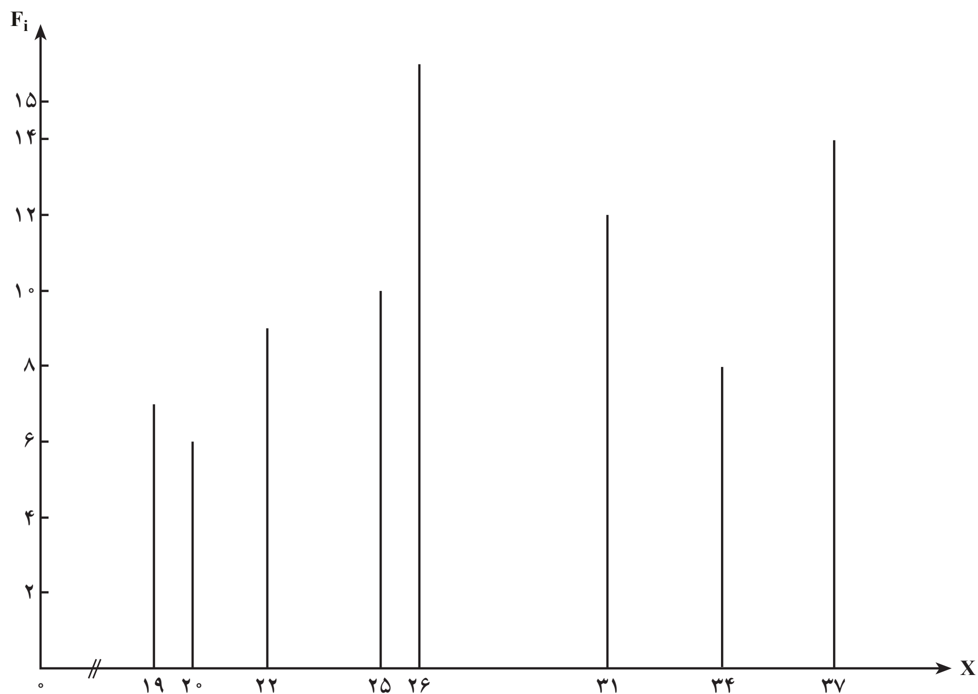
شکل ۲ــ نمودار چندگوش توزیع فراوانیهای تعداد اعضای خانوار

نمودار میله‌ای و چندگوش توزیع فراوانیها را برای مثال ۲ (سن کارگران کارخانهٔ نساجی) رسم می‌کنیم (شکلهای ۳ و ۴) :

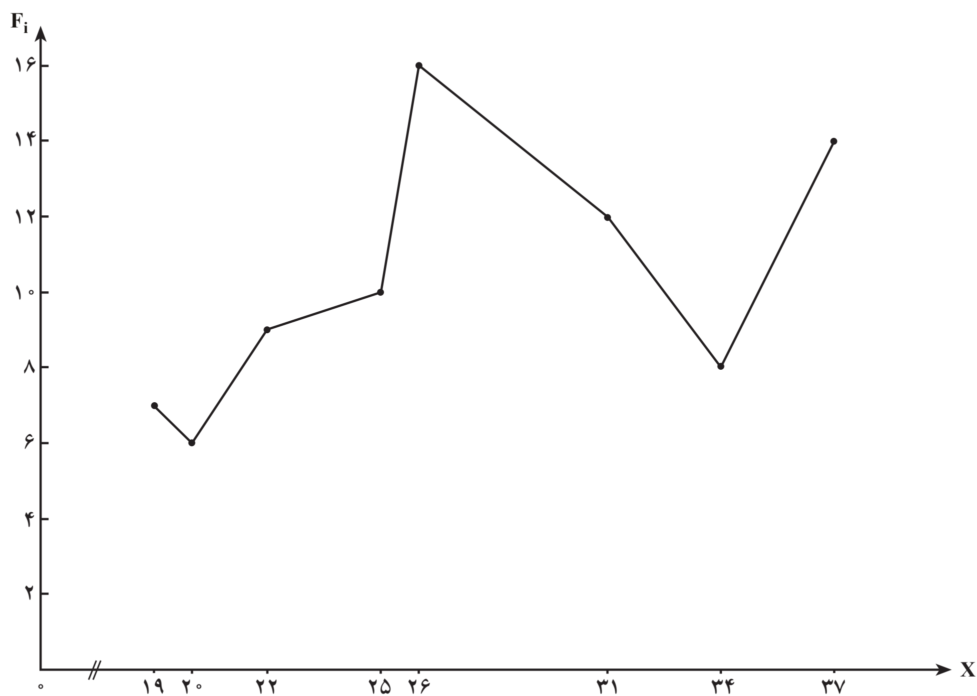
فراوانیهای تجمعی[1] (انباشته)

فراوانیهای تجمعی نیز مانند فراوانیهای مطلق و نسبی، حجم گروهی از عناصر جامعه را بیان

۱ــ Cumulative Frequency



شکل ۳ــ نمودار میله‌ای سن کارگران کارخانه نساجی



شکل ۴ــ نمودار چندگوش سن کارگران کارخانه نساجی

می‌کند که به‌عنوان مشخصهٔ صفت متغیّر، مورد استفاده قرار می‌گیرد و به‌صورت زیر تعریف می‌شود :

فراوانی تجمعی هر گروه از صفت متغیّر در جدول توزیع فراوانی، برابر است با فراوانی مطلق همان گروه به‌علاوهٔ فراوانیهای مطلق گروههای ماقبل آن. فراوانی تجمعی(انباشته) مطلق را با نماد FC_i یا CF_i نشان می‌دهند.

اگر فراوانیهای مطلق را از ابتدا در جدول برای گروههای متوالی جمع کنیم، نهایتاً آخرین فراوانی انباشته (تجمعی) برابر با N یعنی حجم جامعه خواهد شد.

فرم کلی محاسبات فراوانی انباشته به‌شکل زیر است (جدول ۷) :

جدول ۷

X_i	F_i	FC_i
$x_۱$	$F_۱$	$FC_۱ = F_۱$
$x_۲$	$F_۲$	$FC_۲ = F_۱ + F_۲$
$x_۳$	$F_۳$	$FC_۳ = F_۱ + F_۲ + F_۳$
x_k	F_k	$FC_k = F_۱ + F_۲ + F_۳ + \quad + F_k = N$
–	N	–

فراوانیهای انباشته را برای مثال ۱ محاسبه می‌کنیم (جدول ۸) :

جدول ۸

X_i	F_i	FC_i
۲	۴	$FC_۱ = ۴$
۳	۵	$FC_۲ = ۴ + ۵ = ۹$
۴	۷	$FC_۳ = ۴ + ۵ + ۷ = ۱۶$
۵	۴	$FC_۴ = ۴ + ۵ + ۷ + ۴ = ۲۰$
–	۲۰	–

ملاحظه می‌شود که آخرین فراوانی انباشته در جدول برابر ۲۰ یعنی حجم جامعه می‌باشد.

همین مفهوم به‌صورت نسبت، به‌نام فراوانیهای انباشته نسبی نیز بیان می‌شود که به‌صورت زیر تعریف می‌گردد :

> نسبت هریک از فراوانیهای تجمعی (انباشته) در جدول توزیع فراوانی را به N (حجم جامعه) فراوانی انباشته نسبی همان گروه صفت متغیّر می‌نامند و آن‌را با نماد cf_i یا fc_i نشان می‌دهند.

$$fc_i = \frac{FC_i}{N} \qquad (۴)$$

از حاصل جمع تجمعی فراوانیهای نسبی در جدول توزیع فراوانی، به همان ترتیبی که فراوانیهای انباشته مطلق را محاسبه نمودیم نیز می‌توان فراوانی انباشته نسبی هریک از گروههای صفت متغیّر را به‌دست آورد. جدول توزیع فراوانیهای انباشته مطلق و نسبی را برای مثال ۲ محاسبه می‌کنیم :

جدول ۹

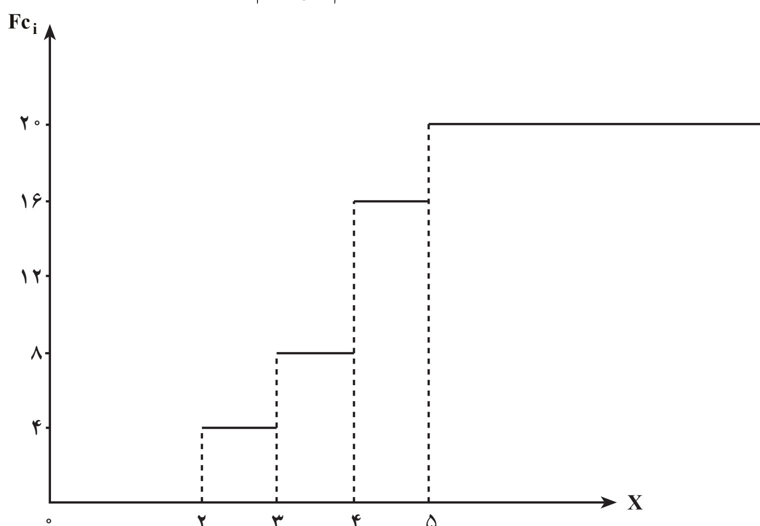
X_i	F_i	f_i	FC_i	fc_i
۱۹	۷	$\frac{۷}{۸۰}$	۷	$\frac{۷}{۸۰}$
۲۰	۶	$\frac{۶}{۸۰}$	۷+۶=۱۳	$\frac{۷}{۸۰}+\frac{۶}{۸۰}=\frac{۱۳}{۸۰}$
۲۲	۹	$\frac{۹}{۸۰}$	۱۳+۹=۲۲	$\frac{۱۳}{۸۰}+\frac{۹}{۸۰}=\frac{۲۲}{۸۰}$
۲۵	۱۰	$\frac{۱۰}{۸۰}$	۲۲+۱۰=۳۲	$\frac{۲۲}{۸۰}+\frac{۱۰}{۸۰}=\frac{۳۲}{۸۰}$
۲۶	۱۵	$\frac{۱۵}{۸۰}$	۳۲+۱۵=۴۷	$\frac{۳۲}{۸۰}+\frac{۱۵}{۸۰}=\frac{۴۷}{۸۰}$
۳۱	۱۲	$\frac{۱۲}{۸۰}$	۴۷+۱۲=۵۹	$\frac{۴۷}{۸۰}+\frac{۱۲}{۸۰}=\frac{۵۹}{۸۰}$
۳۴	۸	$\frac{۸}{۸۰}$	۵۹+۸=۶۷	$\frac{۵۹}{۸۰}+\frac{۸}{۸۰}=\frac{۶۷}{۸۰}$
۳۷	۱۳	$\frac{۱۳}{۸۰}$	۶۷+۱۳=۸۰	$\frac{۶۷}{۸۰}+\frac{۱۳}{۸۰}=\frac{۸۰}{۸۰}=۱$
−	۸۰	۱	−	−

فراوانی‌انباشته برای بیان تعداد مقادیر صفت x معیّن که از یک مقدار مشخص تجاوز نکنند، به‌کار می‌رود. در مثال ۲ اگر بخواهیم بدانیم مجموع کارگرانی که سن آنها از ۲۶ سال تجاوز نمی‌کند،

چه تعدادی است، کافی است فراوانی انباشته ردیف ۲۶ سالگی را مشاهده کرد و آن‌را به‌دست آورد، که در اینجا این رقم برابر ۴۷ می‌باشد. همچنین، اگر بخواهیم نسبت کارگرانی را که سن آنها از ۲۶ سالگی تجاوز نمی‌کند، بدانیم،کافی است، فراوانی‌انباشته‌نسبی سن ۲۶ سالگی را در جدول قرائت نماییم. این نسبت در جدول برابر $\frac{۴۷}{۸۰}$ می‌باشد.

توزیع فراوانیهای انباشته را نیز می‌توان به‌صورت نمودار بیان نمود. برای این‌کار روی محور طولها، مقادیر X و روی محور عرضها، فراوانیهای انباشته‌مطلق یا نسبی را قرار می‌دهیم و برای هر زوج X_i و FC_i یا fc_i یک نقطه در صفحه مختصات به‌دست می‌آوریم و سپس با وصل کردن این نقاط به شکلی دست می‌یابیم که به آن «نمودار توزیع فراوانیهای تجمعی» (کومولات[1]) گویند.

نمودار فراوانیهای انباشته را برای مثال ۱ رسم می‌کنیم (شکل ۵) :



شکل ۵ــ نمودار فراوانیهای تجمعی تعداد اعضای خانوار

نمودار فراوانیهای انباشته را برای مثال ۲ رسم می‌کنیم (شکل ۶) :

طریقهٔ عملی گروه‌بندی نتایج مشاهدات (داده‌های آماری)

در مطالعهٔ جامعه‌های با حجم کوچک، پردازش داده‌ها به‌صورت دستی انجام می‌گیرد. ولی در مطالعهٔ جامعه‌های با حجم بزرگ، از پردازش ماشینی یا رایانه استفاده می‌شود.

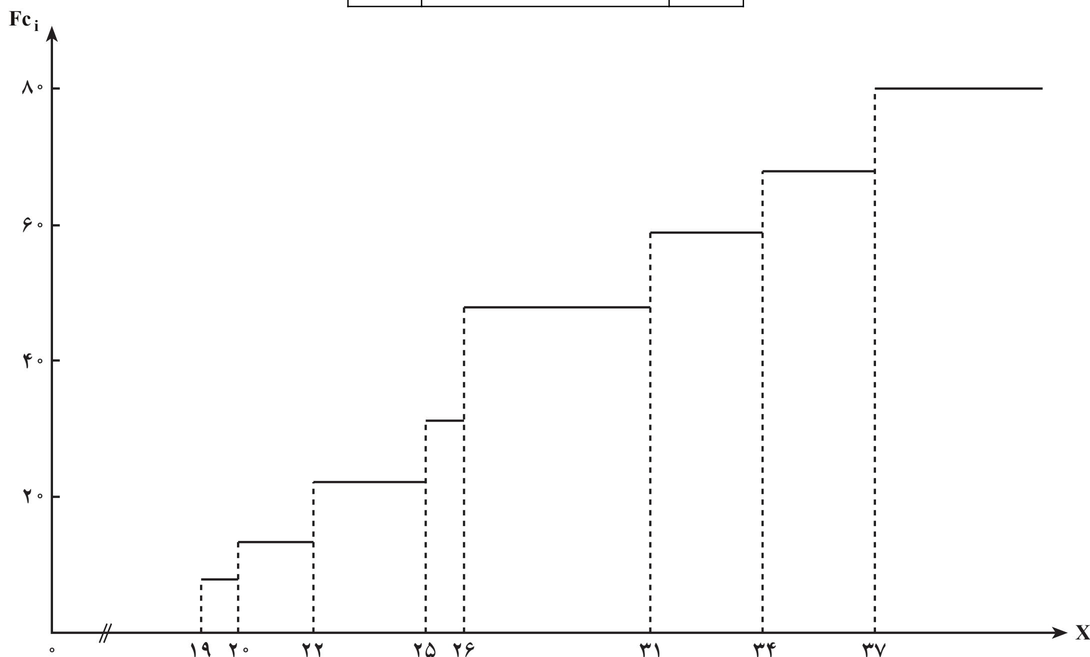
ما در اینجا به‌طور خلاصه فقط دربارهٔ پردازش دستی در مرحلهٔ گروه‌بندی بحث می‌کنیم.

۱ــ Cumulat

ـــ **روش چوب‌خط:** در روش چوب‌خط، جدولها به‌صورت زیر تشکیل می‌گردد :

جدول ۱۰

X	ستون چوب‌خط	F_i
x_1	//	F_1
x_2	///	F_2
x_k	/	F_k
–	–	N



شکل ۶ـــ نمودار فراوانیهای انباشته سن کارگران کارخانه نساجی

در ستون اول جدول، مقادیر صفت متغیّر را به‌طور صعودی از کوچک به بزرگ قرار می‌دهیم. در ستون دوم، برای هریک از مشاهدات در گروه X مربوط، یک چوب‌خط رسم می‌کنیم. این‌کار تا انتقال همهٔ مشاهدات به‌صورت چوب‌خط ادامه می‌یابد. سپس تعداد چوب‌خطها را در هر گروه، شمارش کرده، نتیجه را به‌صورت فراوانی آن گروه در ستون سوم قرار می‌دهیم. در نهایت جمع فراوانیها برابر کل مشاهدات یعنی N خواهد شد.

اگر بخواهیم جدول فراوانیهای فاصله‌ای[1] داشته باشیم، نحوهٔ عملیات، مشابه جدول ۱۰ خواهد بود با این تفاوت که در ستون اول، مقادیر صفت متغیّر را به‌صورت فاصله‌ها قرار می‌دهیم و سپس تک‌تک مشاهدات را در برابر فاصله مربوط با چوب‌خط رسم نموده، از شمارش آنها، فراوانیهای مطلق هر گروه را به‌دست خواهیم آورد که در ستون سوم نوشته می‌شود. به‌این‌ترتیب، جدول توزیع فراوانیهای فاصله‌ای برای صفت متغیّر به‌دست می‌آید (جدول ۱۱) :

جدول ۱۱

فاصله‌های صفت متغیّر	ستون چوب‌خط	فراوانیها
$x_۱ - x_۲$ $x_۲ - x_۳$ $x_۳ - x_۴$ $x_k - x_{k+۱}$	/// ~~////~~ ~~/////~~ //	$F_۱$ $F_۲$ $F_۳$ F_k
–	–	N

ــ مراحل تشکیل جدول توزیع فراوانیهای فاصله‌ای: مراحل تنظیم جدول توزیع فراوانیهای فاصله‌ای را به‌صورت زیر بیان می‌کنیم :

۱ــ پس از انجام مرحلهٔ مشاهده بزرگترین اندازهٔ صفت متغیّر (X_{max}) و کوچکترین اندازهٔ آن را (X_{min}) مشخص کرده، تفاضل آنها را محاسبه می‌کنیم. یعنی

$$X_{max} - X_{min} = R \qquad (۵)$$

این تفاضل را که دامنهٔ تغییرات[2] نامیده می‌شود با R نشان می‌دهند.

۲ــ تعیین تعداد طبقات جدول ــ با توجه به حجم داده‌ها (N) تعداد طبقات را به دلخواه بین ۵ تا ۲۵ طبقه درنظر می‌گیرند. یکی از روشهای کلاسیک برای تعیین تعداد طبقات استفاده از فرمول استروجس است که به‌صورت زیر می‌باشد :

۱ــ Interval Frequency   ۲ــ Range = دامنه

$$K = ۱ + ۳ / ۳۲ \log N \qquad (۶)$$

۳ــ از تقسیم دامنه تغییرات بر تعداد طبقات، فاصلهٔ طبقات به‌دست می‌آید که آن‌را با نماد I[1] نشان خواهیم داد.

$$I = \frac{R}{K} \qquad (۷)$$

۴ــ تعیین فراوانیهای مطلق ــ در این مرحله با کمک روش چوب‌خط، تعداد اعضایی را که اندازهٔ آنها مربوط به هر طبقه جدول است، مشخص می‌کنیم.

با یک مثال، هم طریقهٔ ساختن جدول توزیع فراوانیهای فاصله‌ای و هم طریقهٔ عملی گروه‌بندی نتایج مشاهدات را نشان می‌دهیم :

مثال ۳: تعداد کارگران یک مؤسّسهٔ تولیدی $N = ۱۰۰$ نفر می‌باشند. به‌منظور مطالعهٔ تغییرپذیری سن آنها، تمامی کارگران را مورد مشاهده قرار داده‌ایم. نتایج مشاهدات برای سن آنها، به‌قرار زیر به‌دست آمده است :

۵۰	۲۳	۲۸	۳۹	۵۳	۴۷	۱۸	۲۱	۲۱	۱۸
۱۹	۱۸	۲۲	۲۵	۲۲	۳۳	۴۴	۳۶	۳۳	۳۲
۴۴	۳۰	۳۲	۳۷	۳۹	۱۸	۳۵	۳۵	۲۴	۳۷
۳۲	۲۸	۳۵	۳۶	۳۲	۲۶	۲۷	۲۴	۳۵	۳۵
۴۹	۱۹	۲۰	۱۹	۵۰	۴۹	۴۴	۴۸	۴۶	۳۰
۴۵	۳۹	۳۰	۱۸	۲۰	۳۷	۳۸	۳۰	۴۳	۴۷
۱۸	۲۶	۵۲	۲۸	۲۸	۵۰	۲۸	۵۳	۲۴	۴۴
۴۳	۴۸	۴۱	۴۱	۴۰	۳۸	۳۸	۳۶	۲۷	۲۶
۳۵	۳۰	۳۲	۴۳	۲۵	۴۷	۳۲	۳۳	۳۲	۴۳
۵۴	۵۰	۴۲	۲۹	۴۱	۴۰	۳۸	۳۶	۳۴	۴۷

اگر نتایج مشاهدات را برحسب مقادیر مختلف سن طبقه‌بندی نماییم، ۳۶ گروه سنی تشکیل خواهد شد، که به‌علت زیادی گروههای صفت X، امکان نتیجه‌گیری نسبت به تغییرپذیری سن مشکل است. از این‌رو، در اینجا مناسب است از روش گروه‌بندی فاصله‌ای استفاده کنیم، یعنی با ادغام کردن

۱ــ Interval = فاصله

چند گروه سنی باهم، فاصله‌های سنی بسازیم. با این‌کار، تعداد گروههای صفت متغیّر کمتر و از طولانی شدن جدول کاسته می‌شود.

برای این‌منظور، ابتدا دامنۀ تغییرات صفت را محاسبه می‌کنیم.

در اینجا بزرگترین اندازۀ صفت (سن) ۵۴ سالگی و کوچکترین اندازۀ آن، ۱۸ سالگی است، بنابراین داریم :

دامنۀ تغییرات سن $۵۴ - ۱۸ = ۳۶$

اگر بخواهیم نتایج مشاهدات را در ۸ گروه طبقه‌بندی کنیم $(k = ۸)$، طول فاصله خواهد شد :

طول فاصله $\frac{۳۶}{۸} = ۴/۵ \simeq ۵$

(در چنین مواردی بهتر است طول فاصله را به عدد صحیح تبدیل کنیم).

بنابراین طول فاصله‌هایی را که در جدول توزیع فراوانی تشکیل می‌دهیم، برابر ۵ سال درنظر می‌گیریم. برای تشکیل اوّلین طبقه در جدول توزیع فراوانی، کوچکترین اندازۀ صفت را (X_{min}) برابر کرانه پایین طبقه اول درنظر گرفته، با توجه به مقدار I (فاصله طبقات)، سایر طبقات جدول را تنظیم می‌کنیم. توضیح اینکه کوچکترین اندازه هرطبقه را، کرانۀ پایین آن طبقه و بزرگترین اندازه آن طبقه را، کرانه بالای آن طبقه نامند.

نتایج این گروه‌بندی در جدول زیر نشان داده شده است (جدول ۱۲) :

جدول ۱۲

طبقات سنی X	چوب‌خط	فراوانی F_i
۱۸ – ۲۲	//// //// ////	۱۵
۲۳ – ۲۷	//// //// /	۱۱
۲۸ – ۳۲	//// //// //// ///	۱۸
۳۳ – ۳۷	//// //// //// //	۱۷
۳۸ – ۴۲	//// //// ///	۱۳
۴۳ – ۴۷	//// //// ////	۱۴
۴۸ – ۵۲	//// ////	۹
۵۳ – ۵۷	///	۳
–	–	۱۰۰

باید توجه داشت که هریک از فاصله‌ها از ۵ سال تشکیل شده‌اند. چون سن، یک صفت ناپیوسته است، بنابراین کرانه بالای هرگروه با کرانه پایین گروه بعدی یک واحد تفاوت دارد، به‌عبارت دیگر، سن ۲۲ سالگی در گروه اول قرار می‌گیرد و سن ۲۳ سالگی که بلافاصله بعد از ۲۲ سالگی قرار دارد، در گروه دوم منظور می‌شود. به‌این ترتیب، با قرائت تک‌تک اعداد که درمرحلهٔ مشاهده به‌دست‌آمده است و انتقال آنها به فاصلهٔ مربوط در جدول، به‌وسیلهٔ چوب‌خط، کلیه اعداد را در ستون دوم وارد می‌کنیم و از شمارش آنها، فراوانیهای هر گروه (هر فاصله) را به‌دست می‌آوریم که حاصل جمع آنها نیز برابر N یعنی ۱۰۰ خواهد بود.

اهمیت فراوانیهای فاصله‌ای و مناسب بودن آن، زمانی آشکار می‌گردد که صفت متغیّر مورد مطالعه «پیوسته» باشد.

در مطالعهٔ صفت متغیّر پیوسته، فاصله‌ها را باید چنان انتخاب کرد که صفت در هریک از آنها تقریباً به‌طور یکنواخت توزیع شده باشد. معمولاً انجام این کار وقتی امکان‌پذیر است که فاصله‌های نامساوی انتخاب شوند. لیکن انتخاب فاصله‌های نامساوی، انجام عملیات بعدی را مشکل‌تر می‌سازد. از این‌رو، در عمل اکثر اوقات، نتایج مشاهدات را به‌وسیلهٔ توزیع فراوانیهای با فاصله‌های یکسان بیان می‌کنند.

در استفاده از فاصله‌های با طولهای متفاوت، تحلیل توزیع فراوانیهای فاصله‌ای مشکل‌تر می‌گردد، زیرا فراوانیهای گروهها نه‌تنها از خصوصیّات صفت متغیّر تبعیّت می‌کند، بلکه از طول فاصلهٔ انتخاب شده نیز تأثیر خواهد پذیرفت.

در اینجا به این نتیجه می‌رسیم که در مطالعهٔ صفت متغیّر که به‌صورت توزیع فاصله‌ای بیان شده باشد، باید از مفهوم جدیدی به‌نام «چگالی فراوانی»[1] استفاده کنیم.

چگالی فراوانی، فراوانی را در واحد اندازه صفت نشان می‌دهد.

یعنی :

$$\text{چگالی فراوانی گروه iام} \quad dF_i = \frac{\text{فراوانی گروه iام}}{\text{فاصله همان‌گروه}} = \frac{F_i}{\Delta x_i} \qquad (8)$$

۱ـ Density of Frequency

$$\text{(۹)} \quad \text{چگالی فراوانی نسبی گروه iام} \quad df_i = \frac{\text{فراوانی نسبی گروه iام}}{\text{فاصله همان گروه}} = \frac{f_i}{\Delta x_i}$$

مثال ۴: فرض کنیم نتایج مشاهدات برای صفت متغیّر، موجودی حسابهای پس‌انداز در یکی از شعب بانک باشد که برحسب فاصله‌ها بیان شده است (جدول ۱۳) :

جدول ۱۳

X (موجودی برحسب هزارتومان)	۷–۱۰	۱۰–۱۵	۱۵–۲۵	۲۵–۴۰	۴۰–۶۰	–
F_i فراوانیها	۳۰	۴۰	۶۰	۶۰	۱۰	۲۰۰

اگرچه در توزیع فوق با فاصلهٔ ۱۰ــ۷، (به‌طول ۳ واحد)، فراوانی ۳۰ و با فاصله ۱۵ــ۱۰ (به‌طول ۵ واحد) فراوانی ۴۰ در تناظر است، اما به‌این معنا نیست که با تغییر صفت از فاصله ۱۰ــ۷ به فاصلهٔ ۱۵ــ۱۰ فراوانیها افزایش یافته، بلکه برعکس در فاصله اول به سهم هر واحد طول فاصله $۱۰ = \frac{۳۰}{۳}$ عنصر جامعه تعلق می‌گیرد، در صورتی‌که در فاصلهٔ دوم، به هر واحد طول فاصله فقط $۸ = \frac{۴۰}{۵}$ عنصر جامعه تعلق می‌گیرد.

چگالی فراوانیها را برای فراوانیهای فاصله‌ای محاسبه کرده، توزیع چگالی فراوانیها را به‌دست می‌آوریم.

جدول ۱۴

فاصله‌های صفت	۷–۱۰	۱۰–۱۵	۱۵–۲۵	۲۵–۴۰	۴۰–۶۰
چگالی فراوانی $dF_i = \frac{F_i}{\Delta x_i}$	$\frac{۳۰}{۳} = ۱۰$	$\frac{۴۰}{۵} = ۸$	$\frac{۶۰}{۱۰} = ۶$	$\frac{۶۰}{۱۵} = ۴$	$\frac{۱۰}{۲۰} = ۰/۵$

ملاحظه می‌شود که چگالی فراوانیها با تغییر صفت متغیّر از فاصله‌ای به فاصله دیگر تغییر کرده، نزول می‌یابد. این نزولی بودن تصادفی نیست. برای اطمینان، کافی است فاصله‌ها را عوض کنیم، باز هم خواهیم دید که تغییرات چگالی فراوانیها برای این صفت متغیّر، نزولی است.

اگر در توزیع چگالی فراوانیها، فاصله‌ها را تغییر دهیم، شکل کلی توزیع تغییر نخواهد کرد. بدین‌جهت، در مطالعهٔ توزیع فراوانیهای فاصله‌ای، با فاصله‌های نامساوی، مناسب است، این توزیع فراوانیها، به توزیع چگالی فراوانیها تبدیل گردد.

مثال ۵: فرض کنیم نتایج مشاهدات برای یک صفت متغیّر کمّی پیوسته X، به‌وسیلهٔ جدول توزیع فراوانیهای نسبی فاصله‌ای بیان شده باشد :

جدول ۱۵

فاصله‌های صفت X	۱۱–۱۵	۱۵–۲۵	۲۵–۲۸	۲۸–۳۵	–
فراوانیهای نسبی f_i	۰/۰۸	۰/۴۰	۰/۲۴	۰/۲۸	$\sum f_i = ۱$

ظاهراً در این توزیع، بین تغییرات فراوانیها با فاصله‌های صفت متغیّر، نظم خاصی دیده نمی‌شود، و بزرگترین فراوانی نسبی متعلق به فاصله ۲۵ـ۱۵ است. حال اگر توزیع فراوانیها را به‌صورت توزیع چگالی فراوانیها، درآوریم، نتیجه به‌شکل جدول زیر به‌دست می‌آید :

جدول ۱۶

فاصله‌های صفت X	۱۱–۱۵	۱۵–۲۵	۲۵–۲۸	۲۸–۳۵	–
فراوانیهای نسبی f_i	۰/۰۸	۰/۴۰	۰/۲۴	۰/۲۸	$\sum f_i = ۱$
چگالی فراوانیهای نسبی df_i	۰/۰۲	۰/۰۴	۰/۰۸	۰/۰۴	–

از توزیع چگالی فراوانیها نتیجه می‌گیریم که حداکثر آن در فاصلهٔ ۲۵ـ۱۵ نیست، بلکه به فاصلهٔ ۲۸ـ۲۵ تعلق دارد. یعنی اکثر عناصر جامعه در این فاصله گروه‌بندی شده‌اند. بنابراین، هرگاه جدول توزیع فراوانی فاصله‌ای، با فاصله‌های نامساوی داشته باشیم، برای درک این مطلب که تراکم فراوانیها در چه فاصله‌ای از صفت متغیّر قرار دارند، باید، ابتدا چگالی فراوانیهای صفت‌متغیّر را محاسبه نموده، با استفاده از آنها، نسبت به تراکم فراوانیها نتیجه‌گیری کنیم.

بیان هندسی توزیع فراوانیهای فاصله‌ای (صفت‌متغیّر پیوسته)

نمودار مستطیلی یا هیستوگرام[1] (بافت‌نگار): برای بیان هندسی توزیع صفت‌پیوسته که با فاصله بیان شده است، از دو نوع نمودار استفاده می‌شود :

ــ نمودار مستطیلی یا هیستوگرام برای توزیع چگالی فراوانیها

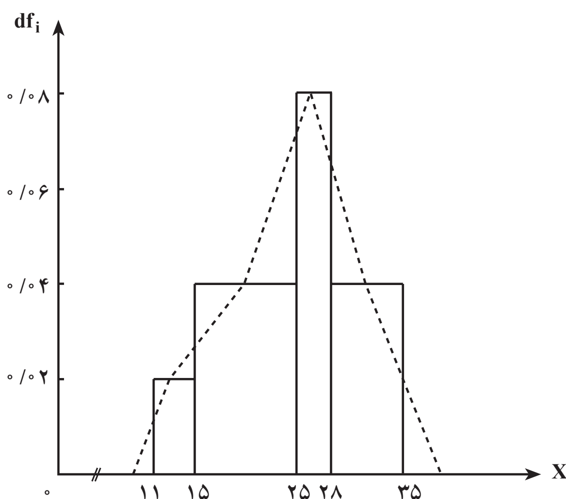
ــ نمودار تجمعی یا اجایو برای توزیع فراوانیهای انباشته

۱ــ Histogram

در ساختن هر دو نوع نمودار، فرض بر این است که چگالی (تراکم) در هریک از فاصله‌ها ثابت باشد.

برای ساختن هیستوگرام، بر روی محور طولها (در دستگاه مختصات قائم) فاصله‌های صفت و بر روی محور عرضها، چگالی فراوانیها را قرار می‌دهند. به‌عبارت دیگر بر روی فاصله‌های صفت‌متغیّر، مستطیلهایی به ارتفاع مساوی با چگالی متناظر با آن فاصله‌ها می‌سازند.

برای جدول ۱۶ در مثال ۵، هیستوگرام توزیع صفت‌متغیّر X را رسم می‌کنیم.



شکل ۷ــ هیستوگرام جدول ۱۶

برای بیان هندسی توزیع صفتی که با فاصله‌ها بیان شده است، می‌توان از نمودار چندگوش نیز استفاده کرد. برای رسم چندگوش توزیع صفت، کافی است روی شکل هیستوگرام، نقاط وسط اضلاع فوقانی مستطیلها را با خطوط مستقیم به‌هم وصل کنیم. از اتصال تمام نقاط وسط مستطیلها به‌هم، نمودار چندگوش به‌دست می‌آید که مساحت زیر آن با محور Xها، برابر با مجموع مساحت مستطیلها، می‌باشد.

در شکل ۷ چندگوش توزیع صفت به‌وسیلهٔ خط‌چین رسم شده است.

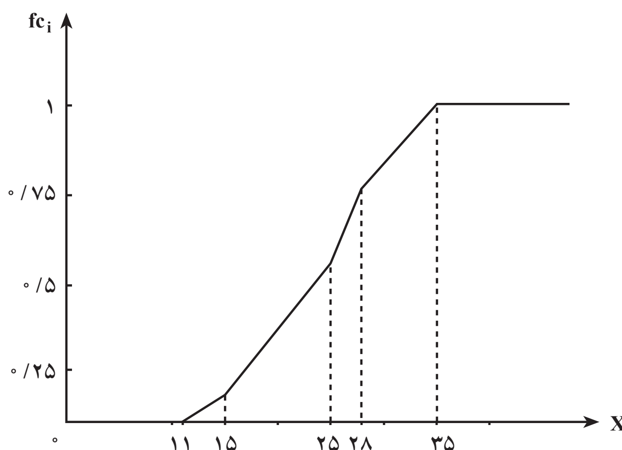
در ساختن نمودار تجمعی یا اجایو با فراوانیهای انباشته، از همان فرض ثابت‌بودن چگالی در فاصله‌های صفت استفاده می‌شود. در صفحه مختصات قائم، روی محور طولها، کرانه‌های هر فاصله را قرار داده، روی محور عرضها، فراوانیهای انباشته را درجه‌بندی می‌کنیم. به‌همان ترتیبی که برای صفت کمّی گسسته، چندگوش فراوانیهای انباشته رسم گردید، عمل می‌کنیم و از به‌دست آوردن نقاط

(باطول X_i و عرض FC_i متناظر)، آنها را به‌وسیلهٔ خطوط مستقیم، به‌هم وصل می‌کنیم. در این‌صورت، شکلی ظاهر می‌شود که نمودار فراوانیهای انباشته نام دارد.

در زیر، جدول توزیع فراوانیهای انباشته نسبی را برای مثال ۵ تشکیل می‌دهیم و با استفاده از آن نمودار تجمعی توزیع صفت را رسم می‌کنیم (جدول ۱۷ و شکل ۸) :

جدول ۱۷

X	۱۱−۱۵	۱۵−۲۵	۲۵−۲۸	۲۸−۳۵	−
f_i	۰/۰۸	۰/۴	۰/۲۴	۰/۲۸	۱
fc_i	۰/۰۸	۰/۴۸	۰/۷۲	۱	−



شکل ۸ ــ نمودار تجمعی (جدول ۱۷) توزیع صفت متغیّر X

می‌توان نشان داد که مساحت هیستوگرام (جمع مساحت مستطیلها) و همین‌طور مساحت زیرخط شکسته چندگوش، وقتی که برای رسم آنها از چگالی فراوانیهای مطلق استفاده شده باشد، همواره مساوی با N است و وقتی که برای رسم آنها از چگالی فراوانیهای نسبی استفاده شده باشد، مساحت هیستوگرام و همین‌طور مساحت زیرخط شکسته چندگوش، مساوی با ۱ است.

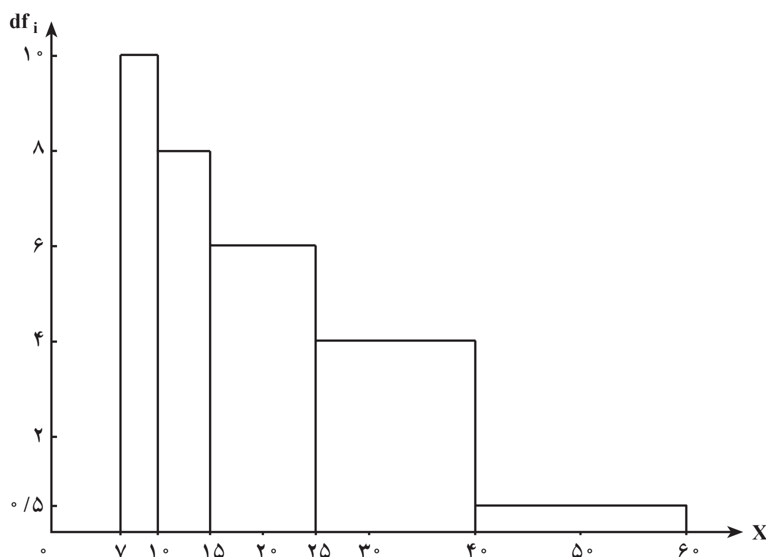
نمودارهای هیستوگرام و تجمعی توزیع صفت را برای مثال ۴ رسم می‌کنیم تا دانش‌آموزان با چگونگی رسم آنها آشنایی بیشتری پیدا کنند.

ابتدا، چگالی فراوانیهای مطلق و فراوانیهای انباشته توزیع صفت را محاسبه می‌کنیم.

جدول ۱۸

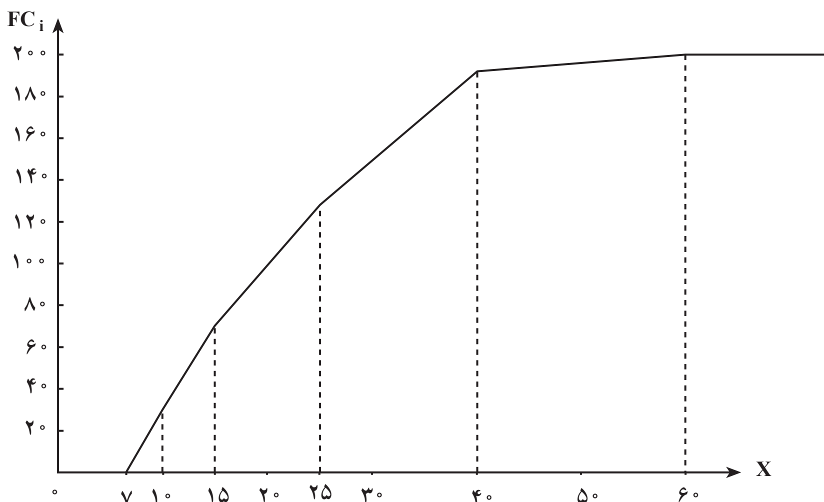
X	۷–۱۰	۱۰–۱۵	۱۵–۲۵	۲۵–۴۰	۴۰–۶۰	–
F_i	۳۰	۴۰	۶۰	۶۰	۱۰	۲۰۰
dF_i	۱۰	۸	۶	۴	۰/۵	–
FC_i	۳۰	۷۰	۱۳۰	۱۹۰	۲۰۰	–

برای رسم هیستوگرام، ابتدا، فاصله‌های صفت را روی محور طولها، در دستگاه مختصات‌قائم وارد کرده، روی محور عرضها، چگالی فراوانیها را قرار می‌دهیم، سپس روی هر فاصله صفت متغیّر، مستطیلی با طول چگالی فراوانی همان فاصله رسم می‌کنیم.



شکل ۹ ــ هیستوگرام توزیع صفت متغیّر X (جدول ۱۸)

برای رسم نمودار تجمعی، اندازه‌های صفت راروی محور طولها و فراوانیهای انباشته را روی محور عرضها در دستگاه مختصات قرار می‌دهیم، سپس با هر مقدار صفت و فراوانی انباشته متناظر آن، یک نقطه روی صفحه مختصات به‌دست می‌آوریم. با وصل نمودن نقاط به‌دست آمده، به‌وسیلهٔ خطوط مستقیم، شکل نمودار تجمعی توزیع صفت متغیّر به‌دست می‌آید.



شکل ۱۰ــ نمودار تجمعی توزیع صفت متغیّر X (مربوط به مثال ۴)

سؤالها و تمرینها

۱ــ لزوم گروه‌بندی نتایج مشاهدات در چیست؟

۲ــ فراوانیهای گروه، چه مفهومی دارد؟

۳ــ توزیع صفت متغیّر، چیست؟

۴ــ فراوانی مطلق، چه مفهومی دارد؟

۵ــ فراوانی نسبی، چه مفهومی دارد؟

۶ــ بین فراوانیهای مطلق چه رابطه‌ای برقرار است؟

۷ــ بین فراوانیهای نسبی چه رابطه‌ای وجود دارد؟

۸ــ توزیع فراوانیهای صفت متغیّر چیست؟ تعریف کنید.

۹ــ فرض کنیم صفت متغیّرکمّی ناپیوسته در جامعه فقط مقادیر ۵، ۱۰، ۱۵، ۲۰ و ۲۵ را به‌ترتیب با فراوانیهای ۵۰، ۱۰۰، ۲۰۰، ۱۰۰و ۵۰ اختیار کرده باشد.

اولاً : توزیع فراوانیهای صفت متغیّر فوق را بیان کنید.

ثانیاً : جدول توزیع فراوانیهای فوق را تشکیل دهید.

۱۰ــ توزیع فراوانیهای نسبی صفت متغیّر ناپیوسته، چگونه به‌دست می‌آید؟

۱۱ــ برای صفت متغیّر در تمرین ۹، فراوانیهای نسبی را محاسبه کنید.

۱۲ــ بین فراوانیهای مطلق و فراوانیهای نسبی چه رابطه‌ای برقرار است؟
۱۳ــ دلیل استفاده از نمودارها برای بیان توزیع صفت متغیّر چیست؟
۱۴ــ نمودار میله‌ای، چگونه ساخته می‌شود؟
۱۵ــ برای توزیع صفت متغیّر در تمرین ۹، نمودار میله‌ای رسم کنید.
۱۶ــ چندگوش توزیع فراوانیها چیست و چگونه ساخته می‌شود؟
۱۷ــ چندگوش توزیع فراوانیها را برای توزیع صفت متغیّر که در جدول زیر داده شده است، رسم کنید.

X	۲	۴	۸	۱۰	۱۲	۱۴	ــ
F_i	۱۲	۲۴	۳۶	۴۸	۲۶	۴	۱۵۰

۱۸ــ توزیع صفت متغیّر X با جدول زیر بیان شده است :

X	۳	۶	۹	۱۲	۱۵	ــ
f_i	۰/۱	۰/۳	۰/۳	۰/۲	۰/۱	۱

نمودار میله‌ای و نمودار چندگوش را برای توزیع فوق رسم کنید.
۱۹ــ فراوانی انباشته (تجمعی) چیست؟ تعریف کنید.
۲۰ــ توزیع فراوانیهای انباشته چیست؟
۲۱ــ توزیع فراوانی صفت متغیّر، با جدول زیر بیان شده است :

X	۲	۴	۶	۸	۱۰	۱۲	ــ
F_i	۲۰	۴۰	۷۰	۳۵	۲۰	۱۵	۲۰۰

جدول توزیع فراوانیهای انباشته را برای صفت متغیر تشکیل دهید.
۲۲ــ فراوانی انباشته نسبی چیست و چگونه محاسبه می‌شود؟
۲۳ــ جدول توزیع فراوانیهای انباشته نسبی را برای تمرین ۲۱ تشکیل دهید.
۲۴ــ توزیع فراوانیهای انباشته به‌صورت هندسی چگونه بیان می‌شود؟
۲۵ــ نمودار تجمعی یا اجایو، برای صفت متغیّر چگونه رسم می‌شود؟
۲۶ــ نمودار تجمعی را برای توزیع صفت در تمرین ۲۱ رسم کنید.

۲۷ــ چرا در گروه‌بندی نتایج مشاهدات صفت متغیّر، از فاصله مقادیر صفت استفاده می‌شود؟

۲۸ــ لزوم تشکیل چنین فاصله‌هایی در چیست؟

۲۹ــ توزیع فراوانیهای فاصله‌ای، چیست؟

۳۰ــ وزن دانش‌آموزان کلاس چهارم دبستان A به‌شرح زیر اندازه‌گیری شده است.

۲۷/۹	۳۱/۸	۳۱/۷	۳۲/۵	۲۸/۵
۳۱/۷	۲۹/۶	۳۲/۵	۳۰/۸	۳۱/۴
۲۸/۹	۲۸/۷	۲۹/۳	۳۰/۴	۳۰/۶
۳۰/۵	۲۸/۶	۲۸/۸	۲۹/۷	۲۹/۳
۳۱/۶	۳۱/۴	۳۳/۵	۳۳/۲	۳۲/۸
۲۸/۲	۲۹/۳	۲۸/۲	۲۷/۸	۳۰/۲
۳۳/۴	۳۲/۸	۳۱/۲	۳۰/۸	۲۹/۸
۳۱/۸	۲۹/۳	۲۹/۹	۲۸/۶	۳۰/۲
۳۴/۴	۳۱/۲	۳۲/۴	۳۴/۳	۳۴/۱
۲۸/۲	۲۷/۸	۳۱/۴	۲۹/۴	۲۸/۷

الف ــ جدول توزیع فراوانیهای گروه‌بندی‌شده وزن دانش‌آموزان را در ۸ گروه تشکیل دهید.

ب ــ جدول توزیع فراوانیهای تجمعی وزن دانش‌آموزان را تشکیل دهید.

ج ــ چند درصد دانش‌آموزان وزنی بیش از ۳۰ کیلوگرم دارند؟

د ــ هیستوگرام وزن دانش‌آموزان را رسم کنید.

۳۱ــ در چه مواردی انتخاب فاصله‌های با طول نامساوی را به فاصله‌های با طول یکسان ترجیح می‌دهید؟

۳۲ــ چگالی فراوانیها چیست و لزوم واردکردن آن در چیست؟

۳۳ــ توزیع فراوانیهای کارگران بر حسب دستمزد روزانه در یک کارگاه ساختمانی به‌صورت جدول زیر بیان شده است :

X دستمزد روزانه (صدتومان)	۱۶ . ۲۵	۲۵ . ۳۰	۳۰ . ۳۵	۳۵ . ۴۰	۴۰ . ۵۰	.
F_i فراوانی	۲۲	۹۸	۶۰	۱۸	۲	۲۰۰

توزیع چگالی فراوانیها را برای دستمزد روزانهٔ کارگران تشکیل دهید.

۳۴ــ دلیل ثبات چگالی فراوانیها در چیست؟

۳۵ــ هیستوگرام یا نمودار مستطیلی چیست؟ در چه مواردی از هیستوگرام استفاده می‌شود؟

۳۶ــ فرضی که در ساختن هیستوگرام باید درنظر گرفت، چیست؟

۳۷ــ برای تمرین ۳۳ هیستوگرام را رسم کنید.

۳۸ــ برای تمرین ۳۳ نمودار تجمعی یا اجایو را رسم کنید.

خودآزمونهای چهارگزینه‌ای

۱ــ کدام رابطه بین فراوانیهای مطلق صحیح است؟

الف ــ $F_i = ۱$.　　ب ــ $F_i = \frac{N}{۲}$.

ج ــ $F_i = ۰$.　　د ــ $F_i = N$.

۲ــ کدام رابطه بین فراوانیهای نسبی صحیح است؟

الف ــ $f_i = ۰$.　　ب ــ $f_i \# ۰/۵$.

ج ــ $f_i = N$.　　د ــ $f_i = ۱$.

۳ــ فراوانی انباشته هر گروه از صفت متغیّر، در جدول توزیع فراوانی برابر است با :

الف ــ فراوانی مطلق گروه مربوط　　ب ــ فراوانی نسبی گروه مربوط

ج ــ فراوانی گروه مربوط به علاوه فراوانیهای ماقبل آن　　د ــ N

۴ــ برای بیان هندسی صفت متغیّر گسسته از کدام نمودارها، استفاده می‌شود؟

الف ــ نمودار فراوانی تجمعی　　ب ــ نمودار میله‌ای و چندگوش

ج ــ هیستوگرام و نمودار مستطیلی　　د ــ نمودار نقطه‌ای

۵ــ برای رسم هیستوگرام روی محورهای مختصّات کدام مؤلفه‌ها قرار داده می‌شود؟

الف ــ فاصله‌های صفت متغیّر و چگالی فراوانی

ب ــ اندازه صفت و فراوانی نسبی　　ج ــ فاصله‌های صفت و فراوانیهای انباشته

د ــ فاصله‌های صفت و فراوانیهای مطلق

۶ــ چگالی فراوانی صفت متغیّر یعنی :

الف ــ فراوانی هر طبقه تقسیم بر N

ب ــ فراوانی هر طبقه تقسیم بر فراوانی نسبی همان طبقه

ج ــ فراوانی هر طبقه تقسیم بر فاصلۀ آن طبقه　　د ــ فراوانی انباشته تقسیم بر N

۷ــ اگر هیستوگرام توزیع صفت با استفاده از چگالی فراوانیهای مطلق، رسم گردد، مساحت مستطیلهای هیستوگرام برابر خواهد شد با :

الف ــ ۱　　ب ــ ۰/۵　　ج ــ N　　د ــ $\frac{N}{۲}$

فصل سوم

مشخص‌کننده‌های مرکزی (تمایل به مرکز)

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند :

۱ــ دلایل به‌کار بردن مشخّصه‌های آماری را توضیح دهد.

۲ــ کاربرد و دلایل استفاده از مشخص کننده‌های مرکزی را بیان کند.

۳ــ مشخّصه‌های مرکزی مهم در توزیع فراوانیهای صفت را به‌صورت جدول، محاسبه نموده، آنها را تعبیر نماید.

۴ــ با استفاده از مشخّصه‌های مرکزی، جامعه‌های مختلف را مقایسه کند.

مشخص‌کننده‌های آماری

در فصل گذشته دربارهٔ توزیع صفت متغیّر درجامعه‌ها بحث شد و نحوهٔ ارایه نتایج مشاهدات، هم با جدول و هم با نمودار، در مورد صفات کمّی گسسته و پیوسته بیان گردید.

ولی برای یک محقق، آگاهی از توزیع صفت در جامعه برحسب تمامی اندازه‌های صفت متغیّر، چندان مفید نیست. اکثر اوقات داشتن اطلاعات اجمالی از خصوصیات صفت در جامعه، برای محقق مفیدتر خواهد بود.

کمیّتهایی که خصوصیات جامعه‌ها را به‌طور اجمال بیان می‌کنند، «مشخّصه‌های عددی» یا «مشخّصه‌های آماری[1]» توزیع صفت، نامیده می‌شوند.

بسته به‌اینکه، چه ویژگیهایی از توزیع صفت، موردنظر باشد، گروههای مختلف مشخّصه‌های آماری، تعریف می‌شوند و مورد استفاده قرار می‌گیرند.

در این فصل دربارهٔ مشخّصه‌های مرکزی (مشخص‌کننده‌های مرکزی)، که نقطهٔ تمرکز توزیع صفت را منعکس می‌کنند، بحث خواهد شد.

مشخّصه‌های مرکزی، مشخّصه‌هایی هستند که مرکزیّت اندازهٔ صفت متغیّر را در جامعه نشان

۱ــ Statistical Charactristic

می‌دهند. مانند : میانگین حسابی (متوسط حسابی)، میانه و نما که جزو این مشخّصه‌ها، به‌حساب می‌آیند.

میانگین حسابی[۱]

یکی از مهمترین مشخّصه‌های مرکزی در آمار، میانگین حسابی یا متوسط حسابی است که به‌صورت زیر تعریف می‌شود :

میانگین حسابی N متغیّر $X_۱$ ، $X_۲$ ، $X_۳$ ، ...، X_N برابر است با :

حاصل جمع متغیّرهای $X_۱$ ، $X_۲$ ، ...، X_N تقسیم بر تعدادشان، یعنی Nکه آن‌را با نماد $\overline{X}$ (بخوانید ایکس‌بار) نشان می‌دهند. (توضیح اینکه غالباً میانگین کل جامعه را با علامت 1 و میانگین مقادیر نمونه را با $\overline{X}$ نشان می‌دهند. در این کتاب برای میانگین، به‌طور کلی از نماد $\overline{X}$ استفاده شده است.)

$$\overline{X} = \frac{X_۱ + X_۲ + X_۳ + \quad + X_N}{N} = \frac{.\ X_i}{N} \qquad (۱)$$

مثال ۱: فرض کنید ۱۰ نفر دانش‌آموز را وزن نموده‌ایم و نتیجه به‌صورت اعداد زیر به‌دست آمده است :

۴۱، ۴۴/۲، ۴۲/۸، ۴۵، ۴۷، ۴۳/۵، ۴۵، ۴۳، ۴۳، ۴۲/۵، ۴۳

می‌خواهیم متوسط وزن این ۱۰ دانش‌آموز را محاسبه کنیم.

اعداد بالا را باهم جمع کرده، نتیجه را بر۱۰ (تعداد آنها) تقسیم می‌کنیم تا متوسط وزن آنها به‌دست آید :

$$\overline{X} = \frac{۴۳ + ۴۲/۵ + ۴۳ + ۴۵ + ۴۳/۵ + ۴۷ + ۴۵ + ۴۲/۸ + ۴۴/۲ + ۴۱}{۱۰} = \frac{۴۳۷}{۱۰} = ۴۳/۷$$

بنابراین متوسط وزن دانش‌آموزان، ۴۳/۷ کیلوگرم می‌باشد.

درصورتی‌که حجم جامعه زیاد باشد و نتایج مشاهدات صفت متغیّر را در جدول توزیع فراوانی بیان کرده باشیم، چون از هرکدام از مقادیر صفت، ممکن است بیش از یک فراوانی داشته باشیم، ابتدا مقادیر صفت را در فراوانی متناظرشان ضرب کرده، نتیجه را در یک ستون در ادامه جدول می‌نویسیم. سپس مقادیر این ستون را باهم جمع کرده، حاصل را بر N تقسیم می‌کنیم. بدین ترتیب، میانگین صفت به‌دست می‌آید، (جدول ۱)، درنتیجه فرمول محاسبه میانگین به‌صورت زیر در می‌آید :

$$\overline{X} = \frac{F_۱X_۱ + F_۲X_۲ + \quad + F_kX_k}{.\ F_i} = \frac{.\ F_iX_i}{N} \qquad (۲)$$

۱ــ Arithmetic Mean

جدول ۱

X_i	F_i	F_iX_i
$X_۱$	$F_۱$	$F_۱X_۱$
$X_۲$	$F_۲$	$F_۲X_۲$
$X_۳$	$F_۳$	$F_۳X_۳$
X_k	F_k	F_kX_k
–	N	. F_iX_i

تمامی محاسبات در جدول ۱ نشان داده شده است.

مثال ۲: توزیع سنی ۵۰ نفر در جدول زیر بیان شده است (جدول ۲). مطلوب‌است میانگین سن برای جامعۀ فوق :

جدول ۲

سن X	۲۰	۲۲	۲۴	۲۶	۲۸	–
فراوانی F_i	۴	۶	۳۰	۷	۳	. $F_i = ۵۰$

برای محاسبۀ میانگین، جدول زیر را تشکیل می‌دهیم (جدول۳) :

جدول ۳

X_i	F_i	F_iX_i
۲۰	۴	۸۰
۲۲	۶	۱۳۲
۲۴	۳۰	۷۲۰
۲۶	۷	۱۸۲
۲۸	۳	۸۴
_	۵۰	۱۱۹۸

$$\overline{X} = \frac{\sum F_iX_i}{N} = \frac{۱۱۹۸}{۵۰} = ۲۳/۹۶$$

بنابراین متوسط سن برای هریک از اعضای این گروه، **۲۳/۹۶** سال می‌باشد. گفتنی است که هرچند سن، یک صفت کمّی گسسته است[۱]، ولی متوسط آن می‌تواند یک عدد غیرصحیح باشد، بنابراین نباید هیچ‌گونه تغییری در آن داده شود.

همانطور که از فرمول میانگین حسابی، دانسته می‌شود، فراوانیهای صفت در آن، مانند وزنه‌هایی هستند که به مقادیر صفت اثر می‌کنند، بدین‌معنی که اگر فراوانی یک مقدار صفت را بزرگ کنیم، متوسط حسابی، به آن مقدار، نزدیک خواهد شد. از این‌رو، میانگین حسابی که طبق فرمول ۲ محاسبه می‌شود، «میانگین حسابی وزنی» نامیده می‌شود.

اگر توزیع صفت متغیّر، با جدول توزیع فراوانیهای نسبی بیان شده باشد، میانگین حسابی از رابطهٔ زیر محاسبه می‌شود:

$$\overline{X} = \sum_{i=۱}^{k} f_iX_i \qquad (۳)$$

خودآزمایی: نشان دهید که دو فرمول زیر معادل هستند:

$$\frac{\sum F_iX_i}{N} = \sum f_iX_i$$

نحوهٔ محاسبه میانگین با فراوانی نسبی در جدول ۴ نشان داده شده است:

جدول ۴

X_i	f_i	f_iX_i
$x_۱$	$f_۱$	$f_۱x_۱$
$x_۲$	$f_۲$	$f_۲x_۲$
x_k	f_k	f_kx_k
–	۱	$\sum f_ix_i$

مثال ۳: مجموعه‌ای از ۲۰۰ سند حسابداری مربوط به خریدهای کوچک، برحسب تعداد

۱ــ سن در تعریف آمار دموگرافیک، عبارت از سالهای تمام شدهٔ عمر انسان؛ بنابراین فقط می‌تواند اعداد صحیح را اختیار کند و یک صفت کمّی گسسته است.

نقصها در این اسناد، گروه‌بندی و نتایج مشاهدات به‌وسیلهٔ جدول توزیع فراوانیهای نسبی زیر، بیان شده است :

X تعداد نقص	۰	۱	۲	۳	۴	–
f_i فراوانی نسبی	۰/۵	۰/۲	۰/۲	۰/۰۵	۰/۰۵	۱

مطلوب است : متوسط تعداد نقصها، برای یک سند.

برای محاسبهٔ متوسط تعداد نقصها، جدول زیر را تشکیل می‌دهیم (جدول ۵) :

جدول ۵

X	f_i	f_iX_i
۰	۰/۵	۰
۱	۰/۲	۰/۲
۲	۰/۲	۰/۴
۳	۰/۰۵	۰/۱۵
۴	۰/۰۵	۰/۲۰
–	۱	۰/۹۵

$\overline{X} = .\ f_iX_i =$ ۰/۹۵

بنابراین متوسط تعداد نقصها برای یک سند، ۰/۹۵ می‌باشد. یعنی به‌طور متوسط در هر سند ۰/۹۵ نقص یا تقریباً یک نقص وجود دارد.

گاهی ممکن است حجم نتایج مشاهدات کوچک باشد. در چنین شرایطی، گروه‌بندی برای صفت انجام نمی‌شود و در این‌صورت میانگین حسابی به‌صورت ساده بیان می‌گردد و همانطورکه قبلاً گفته شد، آن‌را با استفاده از فرمول ۱ محاسبه می‌کنند.

مثال ۴: طول سابقهٔ خدمت ۱۰ نفر مراجعه‌کننده برای اشتغال دریک مؤسّسه (برحسب سال) به‌قرار زیر است :

۱۱، ۷، ۱۰، ۵، ۶، ۸، ۷، ۱۰، ۴، ۸

متوسط حسابی سابقهٔ خدمت را برای یک مراجعه کننده، محاسبه کنید.

دراین حالت به‌علت اینکه نتایج مشاهدات، گروه‌بندی نشده است، میانگین را طبق فرمول ۱

محاسبه می‌کنیم :

$$\overline{X}=\frac{۸+۴+۱۰+۷+۸+۶+۵+۱۰+۷+۱۱}{۱۰}=\frac{۷۶}{۱۰}=۷/۶$$

بنابراین، متوسط سابقه خدمت برای یک مراجعه کننده ۷/۶ سال می‌باشد.

در بسیاری موارد، فراوانیها در توزیع صفت، برحسب فاصله‌های مقادیر صفت بیان می‌شوند. در این‌حالت توزیع صفت را «توزیع فاصله‌ای» می‌نامند.

در محاسبهٔ میانگین برای توزیعهای فاصله‌ای، ابتدا وسط فاصله‌ها را به‌دست آورده، فرض می‌کنند که این اعداد، خود مقادیر مشاهده شده صفت هستند. (یعنی صفت متغیّر را به‌صورت گسسته درنظر می‌گیرند) و سپس میانگین این توزیع جدید را طبق فرمول ۲ یا ۳ محاسبه می‌کنند. روش محاسبهٔ میانگین در چنین حالتی در جدول ۶ نشان داده شده است.

جدول ۶

فاصله گروه‌ها	F_i	X_i	F_iX_i
$X_۱ - X_۲$	$F_۱$	$X_۲$	$F_۱X_۲$
$X_۲ - X_۳$	$F_۲$	$X_۲$	$F_۲X_۲$
$X_k - X_{k+۱}$	F_k	X_k	F_kX_k
−	N	−	. F_iX_i

که در اینجا X_i ها، نقاط میانی فاصله گروه‌هایند و از میانگین کرانهٔ پایین و کرانهٔ بالای هر گروه به‌دست می‌آیند، مثلاً $X_۲$ از $X_۲=\frac{X_۱+X_۲}{۲}$ به‌دست می‌آید.

مثال ۵: مزد و حقوق ماهانهٔ کارکنان یک مؤسسه خدماتی به‌صورت جدول صفحه بعد گروه‌بندی شده است (جدول ۷).

مطلوب است : متوسط مزد و حقوق ماهانهٔ کارکنان این مؤسّسه :

جدول ۷

فراوانی F_i	فاصله طبقات X مزد و حقوق ماهانه برحسب هزار تومان
۴	۴۰–۵۰
۷	۵۰–۶۰
۱۲	۶۰–۷۰
۱۵	۷۰–۸۰
۸	۸۰–۹۰
۴	۹۰–۱۰۰
۵۰	–

برای محاسبهٔ میانگین مزد و حقوق ماهانه، ابتدا وسط فاصله‌ها یعنی X_i هرطبقه را محاسبه کرده، در ستون سوم جدول ۸ می‌نویسیم و سپس محاسبات را طبق فرمول ۲ انجام می‌دهیم :

جدول ۸

F_iX_i	X_i وسط فاصله‌ها	F_i	X فاصله طبقات
۱۸۰	۴۵	۴	۴۰–۵۰
۳۸۵	۵۵	۷	۵۰–۶۰
۷۸۰	۶۵	۱۲	۶۰–۷۰
۱۱۲۵	۷۵	۱۵	۷۰–۸۰
۶۸۰	۸۵	۸	۸۰–۹۰
۳۸۰	۹۵	۴	۹۰–۱۰۰
۳۵۳۰	–	۵۰	–

$$\overline{X} = \frac{\sum F_iX_i}{N} = \frac{۳۵۳۰}{۵۰} = ۷۰/۶$$

بنابراین، متوسط مزد و حقوق ماهانهٔ یک نفر در مؤسسه فوق، ۷۰۶۰۰ تومان است.

مثال ۶: روزهای غیبت کارگران در یک کارگاه ساختمانی به‌صورت جدول زیر است :

X روزهای‌غیبت	۴ – ۰	۹ – ۵	۱۴ – ۱۰	۱۹ – ۱۵	۲۴ – ۲۰	۲۹ – ۲۵	–
F_i فراوانی	۵	۸	۱۰	۶	۷	۴	۴۰

متوسط روزهای غیبت کارگران را محاسبه کنید.

برای محاسبهٔ میانگین روزهای غیبت کارگران جدول زیر را تشکیل می‌دهیم (جدول ۹) :

جدول ۹

X فاصله طبقات	F_i	X. وسط فاصله‌ها	F_iX_i
۴ – ۰	۵	۲	۱۰
۹ – ۵	۸	۷	۵۶
۱۴ – ۱۰	۱۰	۱۲	۱۲۰
۱۹ – ۱۵	۶	۱۷	۱۰۲
۲۴ – ۲۰	۷	۲۲	۱۵۴
۲۹ – ۲۵	۴	۲۷	۱۰۸
–	۴۰	–	۵۵۰

$$\overline{X} = \frac{.\ F_iX_i}{N} = \frac{۵۵۰}{۴۰} = ۱۳/۷۵$$ روز

میانگین حسابی روزهای غیبت کارگران ۱۳/۷۵ روز می‌باشد.

ــ بعضی خواص ریاضی میانگین حسابی

۱ــ میانگین حسابی N کمیّت ثابت برابر است با خود آن کمیّت :

فرض کنید $X = a$ باشد، در این‌صورت

$$\overline{X} = \frac{.\ F_ia}{.\ F_i} = \frac{a.\ F_i}{.\ F_i} = a$$

۲ــ اگر فراوانیهای مقادیر صفت را چند برابر بزرگ یا کوچک کنیم، میانگین حسابی تغییری نمی‌کند. برای مثال اگر تمام فراوانیها را K برابر کوچک کنیم، خواهیم داشت :

$$\overline{X}=\frac{.\ \frac{F_i}{k}X_i}{.\ \frac{F_i}{k}}=\frac{\frac{۱}{k}.\ F_iX_i}{\frac{۱}{k}.\ F_i}=\frac{.\ F_iX_i}{.\ F_i}$$

۳ـ حاصل جمع انحرافهای (تفاضلهای) مقادیر صفت از میانگین خود، همواره مساوی صفر است، یعنی :

$$\boxed{.\ (X_i-\overline{X})=۰}\Rightarrow .\ X_i-.\ \overline{X}=.\ X_i-N\overline{X}=.\ X_i-.\ X_i=۰$$

و به‌همین ترتیب

$$\boxed{.\ F_i(X_i-\overline{X})=۰}$$

۴ـ اگر از تمامی مقادیر صفت، یک عدد ثابت کم شود و یا یک عدد ثابت به آنها اضافه گردد، از میانگین این کمیّتهای جدید نیز ، همان مقدار ثابت کاسته، یا به آن افزوده خواهد شد.

$$(\overline{X+a})=\overline{X}+a$$
$$(\overline{X-a})=\overline{X}-a$$

۵ـ اگر تمامی مقادیر صفت را، b برابر بزرگ یا کوچک کنیم، متوسط حسابی آنها نیز b برابر بزرگ یا کوچک خواهد شد :

$$(\overline{bX})=b\overline{X}$$
$$(\overline{\frac{X}{b}})=\frac{۱}{b}\overline{X}\qquad b\neq ۰$$

محاسبه میانگین به روش کوتاه: خواص بالا می‌توانند برای ساده کردن محاسبهٔ میانگین به‌کار روند که «روش کوتاه» نامیده می‌شوند. در کاربرد این خواص، فرمول محاسبهٔ میانگین را به‌صورت زیر می‌توان بیان کرد :

$$\overline{X}=b\overline{Y}+a \qquad (۴)$$

که در آن

$$\overline{Y}=\frac{.\ F_iY_i}{N}$$

و

$$Y_i=\frac{X_i-a}{b}$$

a و b اعداد دلخواه هستند (درجدول توزیع فراوانیها با فاصله‌های مساوی، برای سهولت کار، a را حدّ وسط یکی از طبقات میانی جدول و b را فاصلهٔ طبقات درنظر می‌گیرند). چون از تمامی X_i ها عدد a کم شده است، بنابراین متوسط کمیّت جدید به اندازهٔ عدد a کوچک خواهد شد (خاصیّت ۴) و چون اعداد $(X_i - a)$ بر عدد b تقسیم شده‌اند، متوسط این اعداد نیز، b برابر کوچکتر خواهد شد (خاصیّت ۵). از این‌رو، متوسط به‌دست آمده برای مقادیر Y_i ها (یعنی $\overline{Y}$) باید در عدد b ضرب و به حاصل، عدد a اضافه شود.

مثال ۷: قد دانش‌آموزان یک کلاس دبیرستانی را اندازه‌گیری کرده‌ایم و نتایج مشاهدات به‌صورت جدول توزیع فراوانی زیر به‌دست آمده است (جدول ۱۰) :

جدول ۱۰

F_i	X فاصله طبقات
۳	۱۴۰–۱۴۵
۴	۱۴۵–۱۵۰
۱۰	۱۵۰–۱۵۵
۱۶	۱۵۵–۱۶۰
۸	۱۶۰–۱۶۵
۶	۱۶۵–۱۷۰
۳	۱۷۰–۱۷۵
۵۰	–

متوسط قد دانش‌آموزان این کلاس را با استفاده از روش کوتاه محاسبه می‌کنیم.

برای این‌منظور، ابتدا وسط فاصله‌ها را به‌دست می‌آوریم (این اعداد در ستون سوم جدول ۱۱ قرار داده شده‌اند).

جدول ۱۱

X_i	F_i	X_i	$X_i - ۱۵۷/۵$	$Y_i = \frac{X_i - ۱۵۷/۵}{۵}$	F_iY_i
۱۴۰–۱۴۵	۳	۱۴۲/۵	–۱۵	–۳	–۹
۱۴۵–۱۵۰	۴	۱۴۷/۵	–۱۰	–۲	–۸
۱۵۰–۱۵۵	۱۰	۱۵۲/۵	–۵	–۱	–۱۰
۱۵۵–۱۶۰	۱۶	۱۵۷/۵	۰	۰	۰
۱۶۰–۱۶۵	۸	۱۶۲/۵	۵	۱	۸
۱۶۵–۱۷۰	۶	۱۶۷/۵	۱۰	۲	۱۲
۱۷۰–۱۷۵	۳	۱۷۲/۵	۱۵	۳	۹
–	۵۰	–	–	–	۲

به‌منظور ساده کردن محاسبات، عدد دلخواه a را، مساوی با ۱۵۷/۵ درنظر می‌گیریم. از تمامی مقادیر وسطی X_iها، عدد دلخواه a را کم کرده، تفاضلها را در ستون ۴ جدول فوق قرار داده‌ایم. سپس تفاضلهای ستون ۴ را بر عدد $b = ۵$ تقسیم کرده، حاصل را در ستون ۵ جدول می‌نویسیم که این اعداد را با Y_i نشان داده‌ایم. حال آنها را در فراوانیهای متناظر ضرب کرده (که حاصل در ستون ۶ جدول نوشته شده است). حاصل جمع را برای محاسبهٔ میانگین Y_iها به‌دست می‌آوریم. بنابراین خواهیم داشت:

$$\overline{Y} = \frac{.\ F_iY_i}{N} = \frac{۲}{۵۰} = ۰/۰۴$$

حال، متوسط صفت X را طبق فرمول ۴ محاسبه می‌کنیم:

$$\overline{X} = b\overline{Y} + a$$

$$= ۵ \times ۰/۰۴ + ۱۵۷/۵ = ۱۵۷/۷$$

ــ مقایسهٔ توزیعهای فراوانی به‌وسیلهٔ میانگین و نتیجه‌گیری نسبت به یکسان بودن آنها: از تعریف میانگین ملاحظه می‌شود که میانگین حسابی صفت متغیّر، تابعی از توزیع صفت است، یعنی هم از مقادیر صفت و هم از فراوانیهای آن مقادیر، تبعیّت می‌کند. بنابراین، اگر دو توزیع (دو جامعه)، دو میانگین حسابی متفاوت داشته باشند، آنگاه دو توزیع متفاوت خواهند بود. ولی اگر میانگینهای دو توزیع یکسان باشند، نمی‌توان نتیجه‌گیری کرد که دو توزیع یکسان هستند، زیرا ممکن است دو توزیع

یکسان باشند و نیز ممکن است یکسان نباشند. به‌عبارت دیگر، دو توزیع یکسان، میانگینهای حسابی یکسانی را مشخص می‌نمایند، ولی دو میانگین حسابی یکسان، دو توزیع یکسان را مشخص نمی‌کنند. به‌عنوان مثال، دو توزیع زیر را می‌توان درنظر گرفت (جداول ۱۲ و ۱۳) :

جدول ۱۲

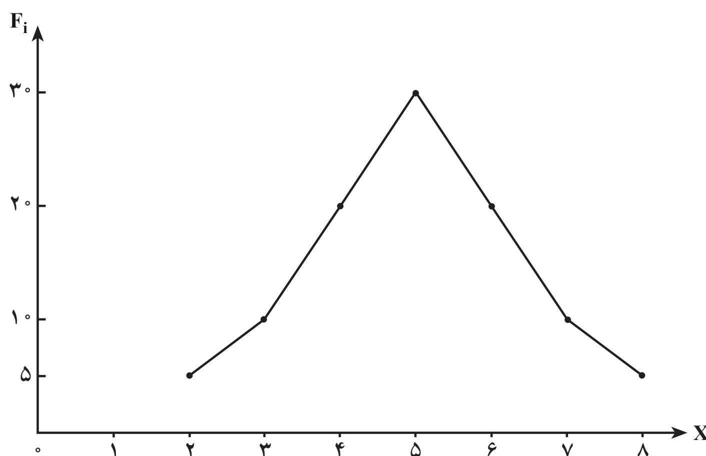
X_i	۲	۳	۴	۵	۶	۷	۸	–
F_i	۵	۱۰	۲۰	۳۰	۲۰	۱۰	۵	۱۰۰

جدول ۱۳

X	۳	۴	۵	۶	۷	–
F_i	۲	۳	۹۰	۳	۲	۱۰۰

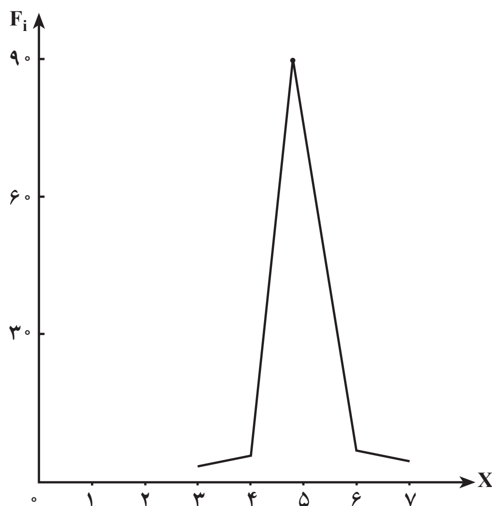
متوسط حسابی برای هر دو توزیع یکسان می‌باشد. $\overline{X}_1 = \overline{X}_2 = 5$

ولی به‌وضوح دیده می‌شود که این دو جدول، دو توزیع متفاوت را بیان می‌کنند. برای روشن شدن این تفاوت، می‌توان دو توزیع را با چندگوش نیز، نمایش داد (شکلهای ۱ و ۲) :



شکل ۱ــ چندگوش توزیع صفت متغیّر جدول ۱۲

بنابراین، همان‌گونه‌که از مثال فوق برمی‌آید، برای پی‌بردن به یکسان بودن توزیعهای صفت متغیّر، همواره نمی‌توان تنها به مشخّصه‌های مرکزی، مانند میانگین حسابی، اکتفا نمود.



شکل ۲ــ چند گوش صفت متغیّر (جدول ۱۳)

میانه[1]

هم‌ردیف با میانگین، به عنوان مشخّصه آماری توزیع صفت، کمیتهای دیگری وجود دارند که در توزیع صفت متغیّر به عنوان مشخّصه مرکزی مورد استفاده قرار می‌گیرند. یکی از چنین مشخّصه‌هایی، میانه می‌باشد که آن را به صورت زیر تعریف می‌کنیم :

میانه، مقدار صفتی است که جامعه را به ۲ گروه با حجم یکسان تقسیم می‌کند.

بر طبق معمول، میانه را با نماد Me نشان می‌دهیم.

از تعریف دقیق میانه برای متغیرهای گسسته در اینجا صرفنظر می‌کنیم[2].

اگر مقادیر صفت متغیّر گسسته را به طور صعودی مرتب کنیم، در صورتی‌که تعداد اعضای جامعه، عدد فرد باشد، یعنی اگر $(N = ۲m + ۱)$ باشد، آنگاه میانه، مقداری است که با شماره $X_{m+۱}$ مشخص می‌شود.

و در صورتی‌که تعداد اعضای جامعه عدد زوج باشد یعنی $(N = ۲m)$ باشد، آنگاه میانه، متوسط حسابی عضو X_m و $X_{m+۱}$ خواهد بود، یعنی :

۱ــ Median

۲ــ به کتابهای آمار ریاضی مراجعه شود.

$$Me = \frac{X_m + X_{m+۱}}{۲}$$

مثلاً اگر مقادیر مشاهدات به‌صورت زیر به دست آمده باشد :

۵،۲،۳،۳،۴،۳،۲،۶،۴،۲،۳

برای تعیین میانه، ابتدا این مقادیر را به‌طور صعودی مرتب می‌کنیم.

۲،۲،۲،۳،۳،۳،۳،۴،۴،۵ ،۶

چون در اینجا $N = ۲m + ۱ = ۱۱$ می‌باشد در نتیجه $m = ۵$ خواهد بود در این صورت عدد وسطی یعنی عضو $X_{m+۱} = X_{۵+۱} = X_۶ = ۳$ میانه می‌باشد.

مثال دیگر ــ اگر مجموعه مشاهدات به‌صورت زیر باشد :

۳،۴،۴،۶،۸،۹

چون در اینجا $N = ۲m = ۶$ و $m = ۳$ ، در نتیجه میانه مقدار صفتی است که با میانگین حسابی X_m و $X_{m+۱}$ مشخص می‌شود. یعنی :

$$Me = \frac{X_m + X_{m+۱}}{۲} = \frac{X_۳ + X_۴}{۲} = \frac{۴ + ۶}{۲} = ۵$$

خواهد بود.

اگر حجم مشاهدات (N)، بزرگ باشد و مقادیر صفت گروه‌بندی شده باشند، دیگر نوشتن تمامی مشاهدات، به صورت دنباله، مشکل و گاهی غیرممکن می‌شود. در چنین حالتی، عضو وسطی را برای توزیع، به وسیلهٔ فراوانیهای انباشته، جستجو می‌کنیم.

مثال ۸: فرض کنید، توزیع صفت متغیّر به‌وسیلهٔ جدول فراوانیهای زیر بیان شده باشد :

جدول ۱۴

X_i	۲	۳	۴	۵	۶	۷	–
F_i	۲	۲	۳	۵	۴	۴	۲۰

برای به‌دست آوردن میانهٔ صفت، ابتدا فراوانیهای انباشته را، محاسبه می‌کنیم :

جدول ۱۵

X_i	۲	۳	۴	۵	۶	۷	–
F_i	۲	۲	۳	۵	۴	۴	۲۰
FC_i	۲	۴	۷	۱۲	۱۶	۲۰	–

حال چنین استدلال می‌کنیم :

حجم مشاهدات $N = ۲۰$ است. باز هم ردیف میانه را مشخص می‌کنیم :

$$Me = \frac{X_m + X_{m+۱}}{۲} = \frac{X_{۱۰} + X_{۱۱}}{۲}$$

یعنی در دنبالهٔ مقادیر صفت، عضو دهم و عضو یازدهم، اعضای وسطی هستند. می‌خواهیم ببینیم این اعضای وسطی (شماره ۱۰ و ۱۱) در کدام گروه قرار دارند. با توجه به فراوانیهای انباشته، مشاهده می‌شود، عضو یکم و دوم در گروه $X = ۲$ هستند.

عضو سوم و چهارم در گروه $X = ۳$ هستند.

عضو پنجم، ششم و هفتم در گروه $X = ۴$ هستند.

عضو هشتم، نهم، دهم، یازدهم و دوازدهم در گروه $X = ۵$ قرار دارند.

بنابراین اعضای وسطی، یعنی عضو دهم و یازدهم، در گروه $X = ۵$ ، یعنی از صفت متغیّر، مقدار $X = ۵$ را دارند و طبق تعریف بالا، خواهیم داشت :

$$Me = \frac{X_{۱۰} + X_{۱۱}}{۲} = \frac{۵ + ۵}{۲} = ۵$$

اگر توزیع صفت، برحسب فاصله‌ها بیان شده باشد (اغلب در حالت صفت متغیّر پیوسته)، پس از محاسبه فراوانیهای انباشته در جدول، فاصله‌ای که $\frac{N}{۲}$ در آن فراوانی انباشته قرار گرفته باشد، آن فاصله را «فاصلهٔ میانه» یا طبقهٔ میانه‌دار می‌نامند و توسط فرمول زیر مقدار میانه را به‌دست می‌آورند :

$$Me = X_i + \frac{\frac{N}{۲} - FC_i}{F_i} \times I \qquad (۵)$$

که در آن :

X_i ــ حد پایین «فاصلهٔ میانه» یا طبقه میانه‌دار

FC_i ــ فراوانی انباشته طبقهٔ قبل از طبقه میانه‌دار

F_i ــ فراوانی مطلق طبقهٔ میانه‌دار

I ــ فاصله طبقهٔ میانه‌دار

می‌باشند.

مثال۹: فرض کنید می‌خواهیم میانهٔ حجم تولید مؤسّسات تولیدکنندهٔ مواد شیمیایی (X) را به‌دست آوریم (جدول ۱۶)، ابتدا ستون فراوانیهای انباشته را تشکیل می‌دهیم :

جدول ۱۶

مقدار محصول X	تعداد مؤسسات F_i	فراوانی انباشته FC_i
۱۳−۱۱	۷	۷
۱۵−۱۳	۹	۱۶
۱۷−۱۵	۱۰	۲۶
۱۹−۱۷	۸	۳۴
۲۱−۱۹	۶	۴۰
−	۴۰	−

سپس نصف حجم مشاهدات را تعیین می‌کنیم : $\frac{N}{۲}=\frac{۴۰}{۲}=۲۰$

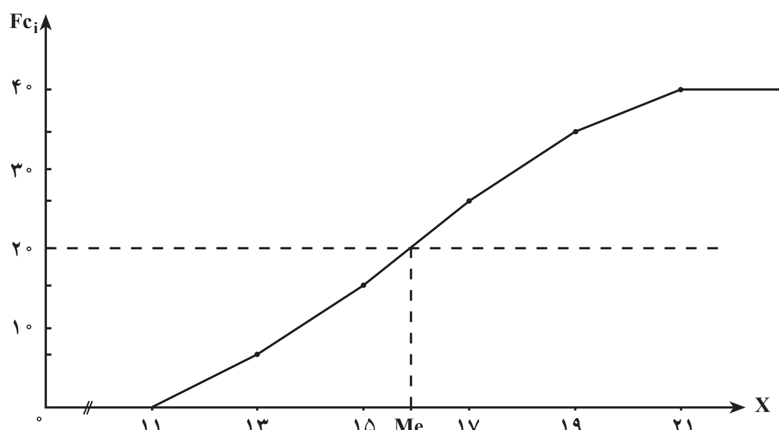
چون نصف حجم جامعه، در فاصله ۱۷ـ۱۵ قرار گرفته، بنابراین، طبقهٔ میانه‌دار، طبقه ۱۷ـ۱۵ می‌باشد، طبق فرمول ۵، میانه را به‌صورت زیر محاسبه می‌کنیم :

$$Me = X_i + \frac{\frac{N}{۲} - FC_i}{F_i} \times I$$

$$= ۱۵ + \frac{۲۰-۱۶}{۱۰} \times ۲$$

$$= ۱۵ + \frac{۸}{۱۰} = ۱۵/۸$$

میانه را می‌توان به‌طور هندسی نیز با استفاده از چندگوش فراوانیهای انباشته، به‌دست آورد (شکل ۳).



شکل ۳ــ تعیین مقدار میانه برای مقدار محصول درمثال (۹)

ــ خواص میانه

۱ــ محاسبهٔ میانه از محاسبهٔ میانگین آسان‌تر است.

۲ــ برای توزیعهای صفت که مقادیر آن در دامنه‌های توزیع، نامعین باشد، میانگین به‌عنوان مقدارمتوسط، نمی‌تواند محاسبه گردد، در صورتی‌که، میانه را برای چنین توزیعهایی، می‌توان محاسبه کرد.

مثلاً، اگر نمایهٔ (لیست) حقوق کارگران یک مؤسسه به‌صورت جدول زیر در دست باشد :

جدول ۱۷

تعداد کارگران F_i	حقوق کارگران (برحسب هزارتومان) X
۱۰	کمتر از ۳۰هزارتومان
۲۰	۳۰ ــ ۴۰
۴۰	۴۰ــ ۵۰
۱۵	۵۰ــ ۶۰
۱۰	۶۰ــ ۷۰
۵	۷۰ هزارتومان به‌بالا
۱۰۰	ــ

برای چنین توزیعی، محاسبهٔ میانگین حسابی، امکان‌پذیر نیست، زیرا باید وسط فاصله‌ها را برای هرگروه محاسبه کرد، که چون گروه اول و گروه آخر، حدود معینی ندارند، بنابراین وسط فاصله آنها را نمی‌توان به‌دست آورد. ولی برای این توزیع، میانه را می‌توان محاسبه نمود، چون میانه با طبقات میانی جدول محاسبه می‌شود و به طبقات اول و آخر جدول کاری ندارد.

۳ــ میانه، حاصل جمع قدرمطلق انحرافات مقادیر صفت را از خودش به حداقل می‌رساند :

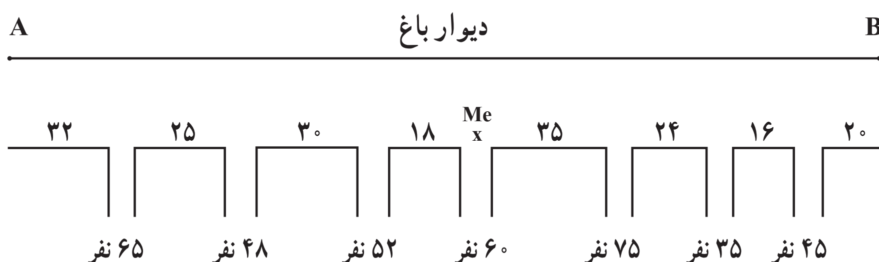
(۶) $|X_i - Me| = \min$.

یا

(۷) $F_i|X_i - Me| = \min$.

از این خاصیت برای نصب یا ایجاد وسایلی که استفاده همگانی دارند، استفاده می‌کنند. مثلاً برای نصب باجه تلفن عمومی در یک معبر و یا تعیین محل مناسب برای ایستگاه اتوبوس در یک خیابان و مواردی از این قبیل، از نقطهٔ میانه این مکانها استفاده می‌شود.

مثال ۱۰: شرکت مخابرات درنظر دارد یک باجه تلفن عمومی در خیابان معینی نصب کند. موقعیّت خیابان و فاصله مکانها و تعداد افراد ساکن در این مکانها در شکل زیر نشان داده شده است :



برای تعیین جای مناسب نصب باجهٔ تلفن، جدول توزیع فراوانی فاصله‌ها را از نقطهٔ A به‌دست آورده، سپس با تشکیل فراوانیهای انباشته، میانهٔ توزیع را تعیین می‌کنیم :

جدول ۱۸

فاصله مکانها تانقطهٔ A X	تعداد نفرات مکانها F_i	فراوانی انباشته FC_i
۳۲	۶۵	۶۵
۵۷	۴۸	۱۱۳
۸۷	۵۲	۱۶۵
۱۰۵	۶۰	۲۲۵
۱۴۰	۷۵	۳۰۰
۱۶۴	۳۵	۳۳۵
۱۸۰	۴۵	۳۸۰
–	۳۸۰	–

نصفِ حجم جامعه $\left(\frac{N}{۲}\right)$ را محاسبه می‌کنیم :

$$\frac{N}{۲}=\frac{۳۸۰}{۲}=۱۹۰$$

با مقایسهٔ $\left(\frac{N}{۲}\right)$، ۱۹۰ با ستون فراوانیهای انباشته، معلوم می‌شود که در فراوانی انباشته ۲۲۵ قرار دارد و گروه X متناظر آن فاصله ۱۰۵ می‌باشد، بنابراین میانهٔ این خیابان نقطهٔ ۱۰۵ متری از نقطهٔ A می‌باشد که در شکل خیابان نشان داده شده است.

چنانچه میانهٔ این خیابان را از نقطهٔ B محاسبه کنیم، میانه، فاصله ۹۵ متری از نقطه B به‌دست خواهد آمد که دقیقاً روی نقطه قبلی خواهد افتاد.

نما (مُد)[1]

یکی‌دیگر از مشخّصه‌های مرکزی «نما» است که به‌صورت زیر تعریف می‌شود :

> «مقدار صفتی که بزرگترین فراوانی یا بزرگترین چگالی فراوانی را در توزیع فراوانی، دارا باشد، «نما» نامیده می‌شود.»

نمای صفت را بعد از این با Mo نشان خواهیم داد.

۱ــ Mode

مثال ۱۱: فرض کنید توزیع صفت متغیّر، به‌وسیلهٔ جدول زیر بیان شده باشد :

جدول ۱۹

X	۲	۳	۴	۵	۶	–
F_i	۱۰	۲۵	۴۰	۲۰	۵	۱۰۰

چون بزرگترین فراوانی در جدول، ۴۰ است و این فراوانی به مقدار ۴ تعلق دارد، بنابراین، نما در این توزیع ۴ خواهد بود، یعنی :

$$Mo = ۴$$

در توزیعهای فاصله‌ای، ابتدا می‌توان، طبقه یا فاصله‌ای را که شامل نما می‌شود، تعیین کرد. این فاصله را «فاصلهٔ نما» یا «طبقهٔ نمادار» می‌نامند. در توزیعهای با فاصله‌های مساوی، فاصلهٔ نما، از روی بزرگترین فراوانی تعیین می‌گردد. ولی در توزیعهای با فاصله‌های نامساوی، فاصلهٔ نما، به‌وسیلهٔ بزرگترین چگالی فراوانی، تعیین می‌شود.

برای به‌دست آوردن نما در توزیعهای با فاصله‌های مساوی، از فرمول زیر استفاده می‌شود :

$$Mo = X_i + \frac{F_i - F_{(i-۱)}}{(F_i - F_{(i-۱)}) + (F_i - F_{(i+۱)})} \times I \qquad (۸)$$

که در آن

X_i ــ حد پایین فاصلهٔ نما

I ــ طول فاصلهٔ نما

F_i ــ فراوانی مطلق متعلق به فاصلهٔ نما (بزرگترین فراوانی در جدول)

$F_{(i-۱)}$ ــ فراوانی متعلق به فاصلهٔ مجاور ماقبل نما

$F_{(i+۱)}$ ــ فراوانی متعلق به فاصلهٔ مجاور بعد از نما

می‌باشند. برای ساده نمودن فرمول ۷ می‌توان از تبدیلهای زیر استفاده نمود :

$$F_i - F_{(i-۱)} = d_۱$$

$$F_i - F_{(i+۱)} = d_۲$$

در این‌صورت فرمول ۸ به‌صورت زیر نوشته می‌شود :

$$Mo = X_i + \frac{d_۱}{d_۱ + d_۲} \times I \qquad (۹)$$

مثال ۱۲: توزیع قد دانش‌آموزان یک کلاس دبیرستانی در جدول ۲۰ آورده شده است. نمای قد دانش‌آموزان را محاسبه نمایید.

جدول ۲۰

X قد دانش‌آموزان	F_i
۱۴۰–۱۴۵	۳
۱۴۵–۱۵۰	۶
۱۵۰–۱۵۵	۱۵
۱۵۵–۱۶۰	۱۸
۱۶۰–۱۶۵	۱۰
۱۶۵–۱۷۰	۸
–	۶۰

چون بزرگترین فراوانی ۱۸ است، گروه متناظر آن، فاصلهٔ ۱۶۰ــ۱۵۵ است. بنابراین مطابق فرمول ۸، نما را محاسبه می‌کنیم:

$$Mo = ۱۵۵ + \frac{۱۸-۱۵}{(۱۸-۱۵)+(۱۸-۱۰)} \times ۵$$

$$= ۱۵۵ + \frac{۱۵}{۱۱} = ۱۵۵ + ۱/۳۵ = ۱۵۶/۳۵$$

نکته: در هر توزیع، به بیش‌از یک میانگین و یک میانه، نمی‌توان دست یافت ولی می‌توان بیش از یک نما به‌دست آورد.

نما را نیز می‌توان با استفاده از شکل توزیع صفت در جامعه، تعیین نمود.

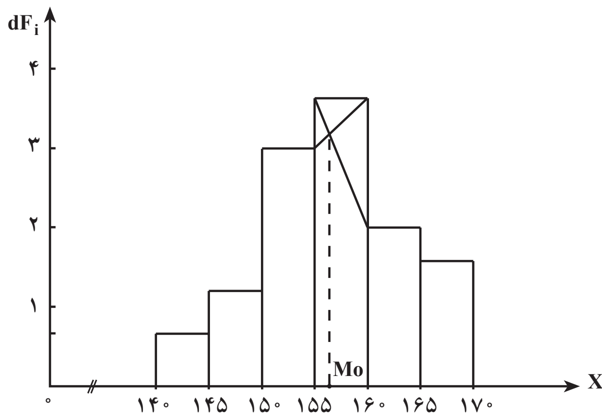
نما را برای صفت متغیّر (قد دانش‌آموزان) درمثال ۱۲ با شکل (هیستوگرام) توزیع به‌دست می‌آوریم.

جدول ۲۱

X	F_i	dF_i
۱۴۰–۱۴۵	۳	۰/۶
۱۴۵–۱۵۰	۶	۱/۲
۱۵۰–۱۵۵	۱۵	۳
۱۵۵–۱۶۰	۱۸	۳/۶
۱۶۰–۱۶۵	۱۰	۲
۱۶۵–۱۷۰	۸	۱/۶
–	۶۰	–

ابتدا چگالی فراوانیها را برای گروههای X در جدول توزیع فراوانی محاسبه می‌کنیم (جدول ۲۱).

سپس هیستوگرام توزیع را رسم می‌کنیم :



شکل ۴ــ تعیین نما با استفاده از هیستوگرام توزیع

در هیستوگرام توزیع، در داخل بلندترین ستون، از ستونهای مجاور، دو خط را مطابق شکل۴ رسم می‌کنیم، سپس از نقطهٔ تقاطع این دو خط، عمودی بر محور طولها، ترسیم می‌کنیم. مقدار صفتی که با طول نقطهٔ عمود بر روی محور طولها، متناظر است، نمای صفت خواهد بود.

در بعضی موارد، نما برای توزیع صفت، یک مشخّصه مرکزی برتر است که بیشترین مقادیر

صفت در حول آن متمرکز می‌گردد. از این‌رو، در توصیف جامعه، در چنین حالاتی به عنوان مشخّصه مرکزی، نما نسبت به میانگین برتری دارد.

مثلاً درمورد تقاضا نسبت به یک اندازه معیّن کفش یا لباس، «نما» نسبت به میانگین یک مشخّصهٔ بهتر است. به‌طور کلی، میانگین در اینجا ارزشی ندارد.

باید متذکر شد که برای توزیعهای کاملاً قرینه، مقادیر میانگین حسابی $(\overline{X})$، میانه (Me) و نما (Mo) برهم منطبق می‌شوند، یعنی :

$$\overline{X} = Me = Mo$$

ولی اگر توزیع قرینه نباشد (البته اگر از حالت قرینگی انحراف بسیار نداشته باشد)، بین این سه مشخّصه‌مرکزی، می‌توان رابطهٔ تقریبی زیر را نوشت :

$$\overline{X} - Mo = ۳(\overline{X} - Me)$$

یا

$$Mo = ۳Me - ۲\overline{X}$$

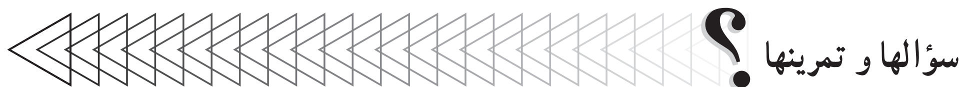
از تساوی تقریبی فوق، می‌توان برای تعیین مقدار یکی از ۳ مشخّصهٔ مرکزی، وقتی که ۲ تای دیگر معلوم باشد، استفاده کرد.

مثلاً برای یک توزیع تقریباً غیرقرینه، اگر میانگین برابر ۴۲ و میانه برابر ۴۰ باشد، با استفاده از رابطهٔ بالا، مقدار نما را محاسبه می‌کنیم :

$$Mo = \overline{X} - ۳(\overline{X} - Me) = ۴۲ - ۳(۴۲ - ۴۰) = ۳۶$$

یا

$$Mo = ۳Me - ۲\overline{X} = ۳ \times ۴۰ - ۲ \times ۴۲ = ۳۶$$



سؤالها و تمرینها

۱ــ مشخّصه‌های آماری چه مشخّصه‌هایی هستند؟ لزوم وارد کردن آنها در آمار چیست؟

۲ــ مشخّصه‌های مرکزی، چه ویژگی از صفت متغیّر را بیان می‌کنند؟

۳ــ میانگین حسابی را تعریف کنید.

۴ــ توزیع میزان فروش یک شرکت بازرگانی درماه گذشته در جدول زیر بیان شده است، میانگین حسابی فروش روزانهٔ این شرکت را محاسبه کنید.

X (میزان فروش برحسب هزار تومان)	۱۰۰ــ۳۰۰	۳۰۰ــ۵۰۰	۵۰۰ــ۷۰۰	۷۰۰ــ۹۰۰	۹۰۰ــ۱۱۰۰	۱۱۰۰ــ۱۳۰۰	ــ
F_i تعداد روز	۳	۷	۱۰	۳	۲	۱	۲۶

۵ــ میانگین حسابی ساده چیست؟ در چه مواردی از آن استفاده می‌کنند؟

۶ــ میانگین حسابی برای توزیعهای فاصله‌ای، چگونه محاسبه می‌شود؟

۷ــ توزیع سن کارگران یک واحد تولیدی، در جدول زیر بیان شده است :

X سن کارگران	۱۸ ــ ۲۲	۲۳ ــ ۲۷	۲۸ ــ ۳۲	۳۳ ــ ۳۷	۳۸ ــ ۴۲	۴۳ ــ ۴۷	ــ
F_i فراوانی	۴	۱۴	۲۰	۱۶	۱۶	۱۰	۸۰

متوسط حسابی سن کارگران را محاسبه کنید.

۸ــ جدول زیر، توزیع خانوارهای یک منطقهٔ مسکونی را برحسب تعداد افراد خانوار نشان می‌دهد :

X تعداد افراد خانوار	۳	۴	۵	۶	۷	۸	ــ
f_i فراوانی نسبی	۰/۰۵	۰/۱۴	۰/۳۰	۰/۲۵	۰/۱۶	۰/۱۰	۱

متوسط تعداد افراد خانوار را در این منطقه محاسبه کنید.

۹ــ خواص میانگین حسابی را بیان کنید. این خواص چه اهمیّتی دارند و کاربردشان چیست؟

۱۰ــ اگر در توزیعهای فاصله‌ای، فاصلهٔ طبقات نامساوی باشد، می‌توان از روش کوتاه، میانگین را محاسبه نمود؟ چرا؟

۱۱ــ میانگین تمرین ۷ را با استفاده از روش کوتاه، محاسبه نمایید.

۱۲ــ میانه را تعریف کنید.

۱۳ــ اگر حجم مشاهدات در جامعه، زوج باشد، میانه را چگونه تعیین می‌کنند؟

۱۴ــ یکی از خواص میانه که لزوم آن‌را در آمار، اجباری می‌سازد، کدام است؟

۱۵ــ اگر توزیع صفت، به وسیلهٔ جدول توزیع فراوانیها، بیان شده باشد، میانه را چگونه محاسبه می‌کنند؟

۱۶ــ تعداد تصادفات اتومبیل در روز، در یک شهر، طی دو ماه گذشته به‌صورت جدول زیر گروه‌بندی شده است :

X تعداد تصادف	۱۰	۱۲	۱۵	۱۶	۱۸	۲۰	–
F_i روز	۳	۱۵	۱۸	۱۴	۷	۳	۶۰

میانه را برای تعداد تصادفات اتومبیل به‌دست آورید.

۱۷ــ مبلغ اضافه‌کار پرداختی دریک شرکت تولیدی در ماه گذشته به‌صورت جدول زیر گروه‌بندی شده است :

X مبلغ اضافه‌کارپرداختی برحسب هزارتومان	۲۰ – ۳۰	۳۰ – ۳۵	۳۵ – ۴۰	۴۰ – ۴۵	۴۵ – ۵۰	–
F_i فراوانی	۸	۱۸	۲۵	۱۲	۷	۷۰

میانه، اضافه کار پرداختی در این شرکت، چقدر است؟

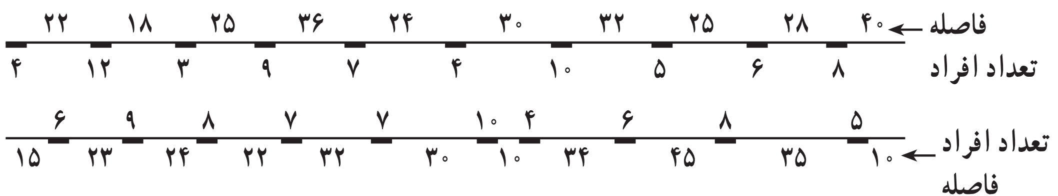
۱۸ــ میزان پارچه‌های تولید شده درکارخانه‌های بافندگی کشور در سال گذشته به‌صورت جدول صفحه بعد گروه‌بندی شده است : (واحد میلیون متر پارچه)

پارچه تولیدشده برحسب میلیون متر X	تعداد کارخانه F_i
۱۰ و کمتر	۵
۱۰–۲۰	۱۷
۲۰–۴۰	۲۸
۴۰–۶۰	۳۵
۶۰–۱۰۰	۲۵
۱۰۰ و بیشتر	۱۰
–	۱۲۰

میانه پارچه تولید شده در کارخانه‌های بافندگی کشور را محاسبه کنید.

۱۹ــ یکی از خواص مهم میانه، که در برنامه‌ریزیها مورد استفاده قرار می‌گیرد، کدام است؟

۲۰ــ فرض کنید شرکت اتوبوسرانی تصمیم دارد در یک خیابان به‌طول ۲۸۰ متر، یک ایستگاه اتوبوس دایر کند، باتوجّه به مشخّصات شکل زیر و با استفاده از خاصیّت میانه، نقطهٔ مناسب محل ایستگاه را مشخص کنید.



(اعداد خارج از معبر فاصله‌ها را و اعداد داخل معبر تعداد افراد ساکن در خانه‌ها را نشان می‌دهد.)

۲۱ــ نما را تعریف کنید.

۲۲ــ نمرات درس ریاضی برای دانش‌آموزان یک کلاس به‌صورت زیر گروه‌بندی شده است :

نمره ریاضی X	۶	۷	۸	۱۰	۱۱	۱۲	۱۴	۱۶	–
فراوانی F_i	۳	۲	۵	۱۲	۶	۱۵	۷	۴	۵۴

نما برای نمرات درس ریاضی، چه نمره‌ای است؟

۲۳ــ نما را برای تمرین ۸، به‌دست آورید.

۲۴ــ نما را برای تمرین ۷ (سن کارگران یک واحد تولیدی) محاسبه کنید.

۲۵ــ نما را برای تمرین ۱۷ (اضافه‌کار پرداختی) محاسبه کنید.

۲۶ــ نما را برای تمرین ۱۶ (تعداد تصادفات اتومبیل) به‌دست آورید.

۲۷ــ نما را برای تمرین ۱۷ با استفاده از شکل هیستوگرام به‌دست آورید.

۲۸ــ در توزیعهای قرینه، چه رابطه‌ای بین مشخّصه‌های مرکزی (میانگین، میانه و نما)، وجود دارد؟

۲۹ــ در توزیعهای غیر قرینه (که انحراف آنها کم است)، چه رابطه‌ای بین میانگین، میانه و نما وجود دارد؟ آن‌را بنویسید.

خودآزمونهای چهارگزینه‌ای

۱ــ کدام‌یک از گزینه‌های زیر، یکی از خواص میانگین را نشان می‌دهد؟

الف ــ $(X_i - \overline{X}) = ۰$.　　ب ــ $(X_i - \overline{X}) = max$.

ج ــ $(X_i - \overline{X}) \# ۰/۵$.　　د ــ $(X_i - \overline{X}) = min$.

۲ــ میانگین ۱۰ عدد برابر ۸ می‌باشد، X_i . کدام است؟

الف ــ ۴۰　　ب ــ ۱۶۰　　ج ــ ۸۰　　د ــ ۱۶

۳ــ اگر به تمامی مقادیر صفت متغیّر، یک عدد ثابت مانند $a = ۵$ اضافه کنیم،

الف ــ میانگین به اندازه ۵ واحد کم می‌شود.

ب ــ میانگین تغییری نمی‌کند.

ج ــ میانگین به اندازهٔ ۵ واحد زیاد می‌شود.

د ــ میانگین دو برابر می‌شود.

۴ــ میانه ۳۲ عضو برابر ۱۶ می‌باشد، اگر فراوانی مطلق طبقهٔ میانه ۶ و فاصله طبقه میانه ۴ و فراوانی انباشته طبقه ماقبل $\frac{N}{۲}$ ، ۱۰ باشد، کرانه پایین طبقه میانه، کدام است؟

الف ــ ۱۰　　ب ــ ۸　　ج ــ ۱۲　　د ــ ۶

۵ــ کدام یک از گزینه‌های زیر یکی از خواصّ میانه می‌باشد؟

الف ــ $(X_i - Me) = ۰$. ب ــ $|X_i - Me| = ۰$.

ج ــ $|X_i - Me| = max$. د ــ $|X_i - Me| = min$.

۶ــ کدام‌یک از گزینه‌های زیر تعریف میانه است؟

الف ــ مقدار صفتی که بیشترین فراوانی را دارد.

ب ــ مقداری از صفت که در وسط توزیع مرتب شده قرار دارد.

ج ــ حاصل‌جمع مقادیر صفت تقسیم بر تعدادشان

د ــ بزرگترین اندازهٔ صفت در جامعه

۷ــ هر توزیع صفت متغیّر

الف ــ فقط یک میانگین و یک نما دارد.

ب ــ فقط یک میانه و یک نما دارد.

ج ــ فقط یک میانگین و یک میانه دارد.

د ــ بیش از یک میانه دارد.

۸ــ نهضت سوادآموزی، زمان پخش برنامه تلویزیونی را از زنان روستایی، نظرخواهی نموده است، کدام مشخّصه مرکزی برای تعیین زمان پخش برنامهٔ نهضت مناسب است؟

الف ــ میانگین ب ــ میانه ج ــ نما د ــ هیچکدام

۹ــ در توزیعهای قرینه، کدام‌یک از روابط زیر صحیح است؟

الف ــ $\overline{X} \# Me \# Mo$ ب ــ $\overline{X} < Me < Mo$

ج ــ $\overline{X} = Me = Mo$ د ــ $\overline{X} = ۳Me - ۲Mo$

۱۰ــ در توزیع‌های تقریباً غیرقرینه کدام رابطه بین سه مشخّصه مرکزی وجود دارد؟

الف ــ $\overline{X} = ۳Me - ۲Mo$ ب ــ $Mo = ۳Me - ۲\overline{X}$

ج ــ $Me = ۳\overline{X} - ۲Mo$ د ــ $Mo = ۳\overline{X} - ۲Me$

فصل چهارم

مشخّصه‌های پراکندگی

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند :

۱ــ دلایل به‌کار بردن مشخّصه‌های پراکندگی را بیان کند.

۲ــ مشخّصه‌های مهم پراکندگی را معرفی کند.

۳ــ نحوهٔ محاسبه واریانس را برای توزیعهای گسسته و پیوسته، به‌کار برد.

۴ــ خواص مشخّصه‌های پراکندگی و کاربرد آنها را بیان کند.

۵ــ برای توزیعهای با دامنهٔ نامعیّن، پراکندگی را با استفاده از چارکها محاسبه کند.

۶ــ مشخّصه‌های نسبی پراکندگی را معرّفی کند.

۷ــ مشخّصه‌های چولگی را برای توزیعهای با دامنهٔ معلوم و نامعیّن محاسبه کند.

۸ــ با مشخّصه‌های پراکندگی بتواند جامعه‌ها را مقایسه کند.

در مطالعه تغییرپذیری صفت متغیّر برای اعضای جامعه، تنها نمی‌توان به متوسّط اندازهٔ صفت، اکتفا نمود، زیرا قبلاً دیدیم که یک میانگین همواره به یک جامعه تعلق ندارد، بلکه دو یا چند توزیع متفاوت، ممکن است میانگینهای یکسان داشته باشند، ولی پراکندگی آنها متفاوت باشد.

برای مشخص‌نمودن پراکندگی صفت متغیّر، می‌توان از انحرافات مقادیر صفت از همدیگر یا انحرافات مقادیر صفت از مشخّصه‌های دیگر مانند مشخّصه‌های مرکزی استفاده کرد.

مشخّصه‌های پراکندگی

طول دامنهٔ تغییرات

در آمار برای اندازه‌گیری پراکندگی، مشخّصه‌های مختلفی مورد استفاده قرار می‌گیرند. یکی از

ساده‌ترین آنها، که به‌آسانی با توزیع صفت متغیّر، محاسبه می‌شود «طول دامنهٔ تغییرات» صفت می‌باشد.

طول دامنهٔ تغییرات از تفاضل بزرگترین اندازهٔ صفت و کوچکترین مقدار آن به‌دست می‌آید.

طول دامنه تغییرات $R = X_{max} - X_{min}$	(۱)

فرض کنیم مقادیر مشاهده شده صفت متغیّر در جامعه به‌قرار زیر به‌دست آمده باشد:

X:۳,۵,۷,۹,۷,۴,۵,۸,۱۱,۱۰

دامنهٔ تغییرات صفت، مطابق فرمول۱ به‌صورت زیر محاسبه می‌شود:

چون بزرگترین اندازهٔ صفت، ۱۱ و کوچکترین اندازهٔ آن، ۳ می‌باشد، بنابراین:

R = ۱۱ − ۳ = ۸

خواهد بود. ولی دامنهٔ تغییرات، به‌عنوان مشخّصه پراکندگی، تصویر صحیحی دربارهٔ چگونگی تراکم مقادیر، حول مقدار متوسط صفت، ارائه نمی‌دهد و فقط از تغییر کوچکترین و بزرگترین مقدار صفت، متأثر می‌شود. از این‌رو، برای ارزیابی نوسانات مقدارصفت نسبت به مقدار متوسط، مشخّصه‌های پراکندگی دیگری مورد استفاده قرار می‌گیرند که ازنظر کلی به دامنهٔ تغییرات ارجحیّت دارند.

متوسط قدرمطلق انحرافات (انحراف متوسّط[1])

اگر بخواهیم نوسانات مقادیر صفت را حول میانگین، با میانگین انحرافات ارزیابی کنیم، یعنی مشخّصه پراکندگی را به‌صورت:

$$\frac{.\ F_i(X_i - \overline{X})}{N}$$

بیان کنیم، چون صورت کسر، مطابق خاصیت ۳ میانگین، همواره مساوی صفر است، دربارهٔ پراکندگی واقعی صفت، چیزی به‌دست نخواهد آمد.

برای برطرف کردن این مشکل، از قدرمطلق انحرافات و یا از مجذور انحرافات استفاده می‌شود. محاسبهٔ میانگین قدرمطلق انحرافات که آن‌را با نمادِ AD نشان می‌دهیم، به‌قرار زیر خواهد بود:

$AD = \dfrac{.\ F_i\left|X_i - \overline{X}\right|}{N}$	(۲)

و اگر فراوانیها برای مقادیر صفت، برابر ۱ باشد، رابطهٔ بالا به‌صورت صفحهٔ بعد درخواهد آمد:

۱ــ Average Deviation =AD

$$AD = \frac{.\ \left|X_i - \overline{X}\right|}{N} \qquad (۳)$$

روش محاسبه آن‌را با ذکر یک مثال بیان می‌کنیم.

مثال ۱: برای توزیع صفت متغیّر که با جدول زیر بیان شده است، متوسط قدرمطلق انحرافات را محاسبه می‌کنیم :

X	۲	۳	۴	۵	۶	–
F_i	۱۰	۲۰	۳۰	۱۶	۴	۸۰

ابتدا میانگین حسابی را برای این توزیع محاسبه می‌کنیم، مقدار آن برابر $\overline{X} = ۳/۸$ است.
بنابراین، طبق رابطهٔ ۲ خواهیم داشت :

$$AD = \frac{.\ F_i\left|X_i - \overline{X}\right|}{N}$$

$$= \frac{۱۰|۲-۳/۸| + ۲۰|۳-۳/۸| + ۳۰|۴-۳/۸| + ۱۶|۵-۳/۸| + ۴|۶-۳/۸|}{۸۰}$$

$$= \frac{۱۸+۱۶+۶+۱۹/۲+۸/۸}{۸۰} = \frac{۶۸}{۸۰} = ۰/۸۵$$

تمام محاسبات می‌تواند در جدول زیر انجام پذیرد (جدول ۱) :

جدول ۱

X	F_i	F_iX_i	$X_i - \overline{X}$	$\left\|X - \overline{X}\right\|$	$F_i\left\|X_i - \overline{X}\right\|$
۲	۱۰	۲۰	۲ – ۳/۸	۱/۸	۱۸
۳	۲۰	۶۰	۳ – ۳/۸	۰/۸	۱۶
۴	۳۰	۱۲۰	۴ – ۳/۸	۰/۲	۶
۵	۱۶	۸۰	۵ – ۳/۸	۱/۲	۱۹/۲
۶	۴	۲۴	۶ – ۳/۸	۲/۲	۸/۸
–	۸۰	۳۰۴	–	–	۶۸

$$\overline{X} = \frac{.\ F_iX_i}{N} = \frac{۳۰۴}{۸۰} = ۳/۸$$

$$AD = \frac{.\ F_i\left|X_i - \overline{X}\right|}{N} = \frac{۶۸}{۸۰} = ۰/۸۵$$

یعنی به‌طور متوسط، هریک از مقادیر صفت، به اندازهٔ ۰/۸۵ واحد از متوسط خود انحراف دارند.

اگر صفت متغیّر با توزیع فاصله‌ای، بیان شده باشد، همانطور که در محاسبهٔ متوسط حسابی، گفته شد، ابتدا وسط فاصله‌ها را تعیین کرده، سپس مانند توزیع گسسته، عمل می‌کنیم.

مثال ۲: متوسط قدرمطلق انحرافات را برای مؤسّسات تولیدکننده مواد شیمیایی در مثال ۹ از فصل سوم، محاسبه می‌کنیم (جدول ۲) :

جدول ۲

X	F_i	X_i	F_iX_i	$\lvert X_i - \overline{X} \rvert$	$F_i \lvert X_i - \overline{X} \rvert$
۱۱–۱۳	۷	۱۲	۸۴	۳/۸۵	۲۶/۹۵
۱۳–۱۵	۹	۱۴	۱۲۶	۱/۸۵	۱۶/۶۵
۱۵–۱۷	۱۰	۱۶	۱۶۰	۰/۱۵	۱/۵۰
۱۷–۱۹	۸	۱۸	۱۴۴	۲/۱۵	۱۷/۲۰
۱۹–۲۱	۶	۲۰	۱۲۰	۴/۱۵	۲۴/۹۰
–	۴۰	–	۶۳۴	–	۸۷/۲۰

$$\overline{X} = \frac{۶۳۴}{۴۰} = ۱۵/۸۵$$

$$AD = \frac{\sum F_i \lvert X_i - \overline{X} \rvert}{N} = \frac{۸۷/۲۰}{۴۰} = ۲/۱۸$$

متوسط قدرمطلق انحرافات، یک اندازهٔ طبیعی برای پراکندگی توزیع می‌باشد، لیکن به‌علت فاقدبودنِ بعضی ویژگیهای ریاضی، از این مشخّصه، کمتر استفاده می‌شود.

متوسط مجذور انحرافات یا واریانس[1] (پراش)

درعمل به‌جای استفاده از علامت قدرمطلق، از مجذور انحرافات استفاده می‌کنند. بنابراین یکی دیگر از معیارهای پراکندگی، متوسط مجذور انحرافات است که آن‌را واریانس می‌نامند و با نماد V(X) نشان داده و با فرمول زیر تعریف می‌شود :

۱ـ Variance

$$V(X)=\frac{F_۱(X_۱-\bar{X})^۲+F_۲(X_۲-\bar{X})^۲+\cdots+F_k(X_k-\bar{X})^۲}{F_۱+F_۲+\cdots+F_k}$$

$$=\frac{\sum F_i(X_i-\bar{X})^۲}{N} \qquad (۴)$$

و اگر مقادیر صفت بدون فراوانی باشند (فراوانی هرکدام یک باشد) :

$$V(X)=\frac{\sum (X_i-\bar{X})^۲}{N} \qquad (۵)$$

محاسبه واریانس را با مثال زیر نشان می‌دهیم :

مثال ۳: برای توزیع صفت متغیّر که با جدول زیر بیان شده است، واریانس را محاسبه می‌کنیم :

X	۲	۳	۴	۵	–
F_i	۱۰	۴۰	۳۰	۲۰	۱۰۰

متوسط حسابی برای این توزیع $\bar{X}=۳/۶$ است، بنابراین با استفاده از فرمول ۴ خواهیم داشت :

$$V(X)=\frac{\sum F_i(X_i-\bar{X})^۲}{N}$$

$$=\frac{۱۰(۲-۳/۶)^۲+۴۰(۳-۳/۶)^۲+۳۰(۴-۳/۶)^۲+۲۰(۵-۳/۶)^۲}{۱۰+۴۰+۳۰+۲۰}$$

$$=۰/۸۴$$

تمامی محاسبات می‌تواند در جدول زیر انجام پذیرد (جدول ۳) :

جدول ۳

X_i	F_i	F_iX_i	$X_i-\bar{X}$	$(X_i-\bar{X})^۲$	$F_i(X_i-\bar{X})^۲$
۲	۱۰	۲۰	۱/۶−	۲/۵۶	۲۵/۶
۳	۴۰	۱۲۰	۰/۶−	۰/۳۶	۱۴/۴
۴	۳۰	۱۲۰	۰/۴	۰/۱۶	۴/۸
۵	۲۰	۱۰۰	۱/۴	۱/۹۶	۳۹/۲
–	۱۰۰	۳۶۰	–	–	۸۴

$$\bar{X}=\frac{۳۶۰}{۱۰۰}=۳/۶$$

$$V(X)=\frac{\sum F_i(X_i-\overline{X})^{۲}}{N}=\frac{۸۴}{۱۰۰}=۰/۸۴$$

در محاسبهٔ واریانس، اغلب به‌علت اینکه متوسط حسابی، با رقمهای اعشاری باید از مقادیر صفت کم شود، به‌توان رساندن این اعداد اعشاری، محاسبه را دشوارتر می‌سازد. از این‌رو، فرمول واریانس را به‌صورت زیر تبدیل می‌کنند :

$$V(X)=\frac{\sum F_i(X_i-\overline{X})^{۲}}{N}=\frac{\sum F_iX_i^{۲}}{N}-\overline{X}^{۲} \qquad (۶)$$

$$V(X)=\frac{\sum F_i(X_i-\overline{X})^{۲}}{N}=\frac{\sum F_iX_i^{۲}-\frac{(\sum F_iX_i)^{۲}}{N}}{N} \qquad (۷)$$

و اگر فراوانی هرکدام از اندازه‌ها مساوی یک باشد، فرمولهای ۶ و ۷ به‌صورت زیر نوشته خواهد شد :

$$V(X)=\frac{\sum (X_i-\overline{X})^{۲}}{N}=\frac{\sum X_i^{۲}}{N}-\overline{X}^{۲} \qquad (۸)$$

$$V(X)=\frac{\sum (X_i-\overline{X})^{۲}}{N}=\frac{\sum X_i^{۲}-\frac{(\sum X_i)^{۲}}{N}}{N} \qquad (۹)$$

فرمولهای ۶ تا ۹ را فرمولهای کاربردی واریانس نامند.

مثال ۴: واریانس را برای توزیع صفت متغیّر که در مثال ۳ بیان شده است، با استفاده از فرمول ۶ محاسبه می‌کنیم (جدول ۴) :

جدول ۴

X_i	F_i	F_iX_i	$F_iX_i^{۲}$
۲	۱۰	۲۰	۴۰
۳	۴۰	۱۲۰	۳۶۰
۴	۳۰	۱۲۰	۴۸۰
۵	۲۰	۱۰۰	۵۰۰
–	۱۰۰	۳۶۰	۱۳۸۰

$$\overline{X}=\frac{۳۶۰}{۱۰۰}=۳/۶$$

$$V(X)=\frac{.\ F_iX_i^{۲}}{N}-\overline{X}^{۲}$$

$$=\frac{۱۳۸۰}{۱۰۰}-(۳/۶)^{۲}=۱۳/۸-۱۲/۹۶=۰/۸۴$$

اگر صفت متغیّر، به‌صورت توزیع فاصله‌ای بیان شده باشد، مانند محاسبهٔ متوسط حسابی، ابتدا، وسط فاصله‌ها را تعیین کرده، سپس مانند توزیع گسسته، عمل می‌کنند.

مثال ۵: واریانس را برای توزیع صفت متغیّر که با جدول ۲ در مثال ۲ بیان شده است، محاسبه می‌کنیم. تمامی محاسبات در جدول ۵ آورده شده است :

جدول ۵

X	F_i	X_i	F_iX_i	$F_iX_i^{۲}$
۱۱–۱۳	۷	۱۲	۸۴	۱۰۰۸
۱۳–۱۵	۹	۱۴	۱۲۶	۱۷۶۴
۱۵–۱۷	۱۰	۱۶	۱۶۰	۲۵۶۰
۱۷–۱۹	۸	۱۸	۱۴۴	۲۵۹۲
۱۹–۲۱	۶	۲۰	۱۲۰	۲۴۰۰
–	۴۰	–	۶۳۴	۱۰۳۲۴

$$V(X)=\frac{.\ F_iX_i^{۲}-\frac{(.\ F_iX_i)^{۲}}{N}}{N}$$

$$=\frac{۱۰۳۲۴-\frac{(۶۳۴)^{۲}}{۴۰}}{۴۰}=\frac{۲۷۵/۱}{۴۰}=۶/۸۷۷۵$$

انحراف معیار[۱]

واریانس، دارای ویژگیهای مهم ریاضی است که در اینجا درباره آنها بحث نمی‌کنیم لیکن به علت اینکه، واریانس، پراکندگی را به صورت مجذور انحرافات بیان می‌کند، برای اینکه متوسط انحراف برای یک عضو مشخص گردد، از واریانس جذر گرفته و آن را انحراف معیار صفت متغیّر می‌نامند و با حرف یونانی . نشان می‌دهند.

۱ــ Standard Deviation

$$= . + \sqrt{V(X)}$$

$$= +\sqrt{\frac{.\ F_i(X_i - \overline{X})^۲}{N}} \qquad (۱۰)$$

با وجود اینکه در آمارنظری، به‌عنوان اندازهٔ پراکندگی، به‌طور اساسی، واریانس مورد استفاده قرار می‌گیرد، ولی انحراف معیار، مناسب‌ترین و متداولترین مشخّصه پراکندگی است که در عمل از آن استفاده می‌شود. از این‌رو، در بحثهای بعدی، پراکندگی را با انحراف معیار، ارزیابی خواهیم کرد.

مثال۶: انحراف معیار را برای مثال ۵ محاسبه می‌کنیم :

$$= . + \sqrt{V(X)} = \sqrt{۶ / ۸۷۷۵} = ۲ / ۶۲۲$$

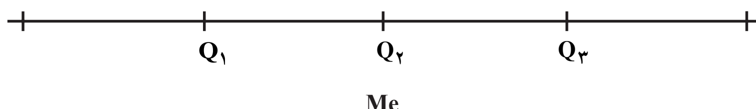
انحراف معیار، نیز مانند متوسط قدرمطلق انحرافات، متوسط انحرافات اندازه‌های صفت را از میانگین صفت نشان می‌دهد.

انحراف چارکی[۱]

به همان دلایلی که میانه به‌عنوان مشخّصه مرکزی تعریف شده است، برای تغییرپذیری جامعه‌هایی که دامنه توزیع صفت در آنها نامشخص باشد، از مشخّصه‌های آماری که از دامنه توزیع متأثر نمی‌باشند، استفاده می‌کنند. چنین مشخّصه‌های آماری، چارکها[۲] هستند که میانه یکی از آنها می‌باشد.

برای تعیین پراکندگی توزیع صفت در چنین حالتی از انحراف چارکی استفاده می‌شود که به‌صورت زیر تعریف می‌گردد :

اگر طول دامنه تغییرات صفت را به چهار قسمت مساوی از نظر حجم تقسیم کنیم، سه عددی که این دامنه را به چهار قسمت افراز می‌کند به ترتیب چارک اول، چارک دوم و چارک سوم نامند و با $Q_۱$ ، $Q_۲$ و $Q_۳$ نشان داده می‌شوند.



متوسط تفاضل چارک دوم و چارک اول با تفاضل چارک سوم و چارک دوم را انحراف چارکی نامند و آن را با نماد QD نشان می‌دهند.

۱ـ Quartile Deviation

۲ـ Quartiles

فرمول آن به صورت زیر بیان می‌شود :

$$QD = \frac{(Q_۲ - Q_۱) + (Q_۳ - Q_۲)}{۲} = \frac{Q_۳ - Q_۱}{۲} \qquad (۱۱)$$

چارکها، مقدار صفت متغیّر در مجموعه مرتب شده مقادیر صفت هستند که ۲۵درصد اعضای جامعه، کوچکتر از $Q_۱$ و ۲۵درصد اعضای جامعه، بین $Q_۱$ و $Q_۲$ و ۲۵درصد اعضای جامعه بین $Q_۲$ و $Q_۳$ و ۲۵درصد بقیه اعضا، بالاتر از $Q_۳$ قرار می‌گیرند.

چارکهای اول و سوم، یعنی $Q_۱$ و $Q_۳$ ، مشابه میانه، طبق فرمولهای زیر محاسبه می‌شوند :

$$Q_۱ = X_i + \frac{\frac{N}{۴} - FC_i}{F_i} \times I \qquad (۱۲)$$

که در آن

X_i ــ نقطهٔ شروع فاصله گروهی از صفت متغیّر است که چارک اول در آن قرار دارد.

I ــ طول فاصله‌ای است که چارک اول در آن قرار دارد.

FC_i ــ فراوانی انباشته گروه قبل از گروه چارک اول می‌باشد.

F_i ــ فراوانی مطلق گروه چارک اول است.

$$Q_۳ = X_i + \frac{\frac{۳N}{۴} - FC_i}{F_i} \times I \qquad (۱۳)$$

که در آن

X_i ــ نقطهٔ شروع فاصله گروهی از صفت متغیّر است که چارک سوم در آن قرار دارد.

I ــ طول فاصله‌ای است که چارک سوم در آن قرار دارد.

FC_i ــ فراوانی انباشته گروه قبل از گروه چارک سوم می‌باشد.

F_i ــ فراوانی مطلق گروه چارک سوم است.

بدیهی است، برای محاسبات چارک اول و چارک سوم، باید ستون فراوانیهای تجمعی را در جدول توزیع صفت، محاسبه کنیم.

مثال ۷: فرض کنید توزیع فراوانی کارگران برحسب مزد و حقوق ماهانه آنها در یک مؤسّسه با جدول زیر بیان شده باشد. می‌خواهیم تغییرپذیری (پراکندگی) مزد و حقوق ماهانه را برای این جامعه محاسبه کنیم :

جدول ۶

مزد و حقوق ماهانه کارگران X	تعداد کارگران F_i	فراوانی انباشته FC_i
تا ۴۰۰۰۰	۱۲	۱۲
۴۵۰۰۰–۴۰۰۰۰	۳۸	۵۰
۵۰۰۰۰–۴۵۰۰۰	۴۴	۹۴
۵۵۰۰۰–۵۰۰۰۰	۵۰	۱۴۴
۶۰۰۰۰–۵۵۰۰۰	۶۲	۲۰۶
۷۰۰۰۰–۶۰۰۰۰	۴۲	۲۴۸
۸۰۰۰۰–۷۰۰۰۰	۲۵	۲۷۳
۱۰۰,۰۰۰–۸۰۰۰۰	۱۴	۲۸۷
۱۰۰,۰۰۰ تومان به بالا	۱۳	۳۰۰
–	۳۰۰	–

ملاحظه می‌شود که مشخّصه‌هایی مانند «طول دامنهٔ تغییرات»، «متوسّط قدرمطلق انحرافات»، واریانس و انحراف معیار به‌عنوان مشخّصه پراکندگی برای این توزیع نمی‌تواند محاسبه گردد. زیرا حدود بالا و پایین جامعه نامشخص است. از این‌رو «انحراف چارکی» را برای این توزیع فراوانیها، محاسبه می‌کنیم. ابتدا ستون فراوانیهای انباشته را محاسبه می‌کنیم که نتیجهٔ آن در ستون سوم جدول ۶ آمده است.

محاسبهٔ چارک اول (Q_1): برای این‌منظور $\frac{N}{4}$ را تعیین می‌کنیم :

$$\frac{N}{4} = \frac{300}{4} = 75$$

با مقایسهٔ مقدار چارک اول (۷۵) با ستون فراوانیهای انباشته جدول، مشاهده می‌شودکه در فراوانی انباشته (۹۴) متعلق به فاصله ۵۰۰۰۰ـ۴۵۰۰۰ قرار دارد، بنابراین چارک‌اول در فاصله ۵۰۰۰۰ـ۴۵۰۰۰ می‌باشد. با استفاده از فرمول ۱۲ مقدار Q_1 را محاسبه می‌کنیم.

در اینجا :

$$X_i = 45000$$

$$I = 50000 - 45000 = 5000$$

$$FC_i = 50$$

$$F_i = 44$$

$$Q_۱ = X_i + \frac{\frac{N}{۴} - FC_i}{F_i} \times I$$

$$= ۴۵۰۰۰ + \frac{۷۵ - ۵۰}{۴۴} \times ۵۰۰۰$$

$$= ۴۵۰۰۰ + \frac{۱۲۵۰۰۰}{۴۴} = ۴۵۰۰۰ + ۲۸۴۱ = ۴۷۸۴۱$$

مشابه روش بالا، چارک سوم ($Q_۳$) را محاسبه می‌کنیم. برای این‌منظور $\frac{۳N}{۴}$ را تعیین می‌کنیم :

$$\frac{۳N}{۴} = \frac{۳ \times ۳۰۰}{۴} = ۲۲۵$$

با $\frac{۳N}{۴}$ ، فاصله‌ای را که شامل چارک سوم است، جستجو می‌کنیم. چون ۲۲۵ در فراوانی انباشته (۲۴۸) که متعلق به فاصلهٔ ۷۰۰۰۰ـ۶۰۰۰۰ است، قرار دارد، از این‌رو، چارک سوم در فاصله ۷۰۰۰۰ـ۶۰۰۰۰ می‌باشد. با استفاده از فرمول ۱۳، چارک سوم را محاسبه می‌کنیم.

در اینجا :

$$X_i = ۶۰۰۰۰$$

$$I = ۷۰۰۰۰ - ۶۰۰۰۰ = ۱۰۰۰۰$$

$$FC_i = ۲۰۶$$

$$F_i = ۴۲$$

$$Q_۳ = X_i + \frac{\frac{۳N}{۴} - FC_i}{F_i} \times I$$

$$= ۶۰۰۰۰ + \frac{۲۲۵ - ۲۰۶}{۴۲} \times ۱۰۰۰۰$$

$$= ۶۰۰۰۰ + \frac{۱۹۰۰۰۰}{۴۲} = ۶۰۰۰۰ + ۴۵۲۴ = ۶۴۵۲۴$$

حال با استفاده از فرمول ۱۱، انحراف چارکی (QD) را محاسبه می‌کنیم :

$$QD = \frac{Q_۳ - Q_۱}{۲}$$

$$= \frac{۶۴۵۲۴ - ۴۷۸۴۱}{۲} = ۸۳۴۱/۵$$

چون تمامی مقادیر صفت متغیّر، بر روی انحراف چارکی، تأثیر نمی‌کنند (حداکثر ۵۰درصد از مقادیر صفت در آن تأثیر دارند)، از این مشخّصه در مواردی که محاسبه انحراف معیار، دشوار یا غیرممکن است، استفاده می‌شود.

مشخّصه‌های نسبی پراکندگی

برای مقایسهٔ تغییرات صفت در دو یا چند جامعه با میانگینهای مختلف، از مشخّصه‌های نسبی پراکندگی، استفاده می‌شود. این مشخّصه‌ها از نسبت مشخّصه‌های پراکندگی به مشخّصه‌های مرکزی صفت، محاسبه می‌شوند.

یکی از مهمترین مشخّصه‌های نسبی پراکندگی، ضریب تغییرات است که به‌صورت زیر تعریف می‌شود :

ضریب تغییرات[1]

نسبت انحراف معیار صفت متغیّر بر متوسط حسابی صفت که برحسب درصد بیان شده باشد، ضریب تغییرات، نامیده می‌شود و آن‌را با C.V نشان می‌دهند.

$$C.V = \frac{\cdot}{\overline{X}} \times ۱۰۰ \qquad (۱۴)$$

که در آن :

. ــ انحراف معیار صفت متغیّر است.

$\overline{X}$ ــ متوسط حسابی صفت متغیّر، می‌باشد.

مثال۸: ضریب تغییرات را برای صفت متغیّر که در مثال ۵ آورده شده است، و برای آن $\overline{X}$=۱۵/۸۵ و ۲/۶۲۲ . = می‌باشند، محاسبه می‌کنیم :

$$C.V = \frac{\cdot}{\overline{X}} \times ۱۰۰ = \frac{۲/۶۲۲}{۱۵/۸۵} \times ۱۰۰ = ۱۶/۵۴$$

یعنی تغییرات صفت، ۱۶/۵درصد اندازه متوسط آن را تشکیل می‌دهد.

ضریب تغییرات نه‌تنها برای بیان تغییرپذیری نسبی به‌کار می‌رود، بلکه به عنوان «مشخّصه همگنی» جامعه نیز مورد استفاده قرار می‌گیرد. هرگاه ضریب تغییرات صفت در جامعه از ۳۳درصد تجاوز نکند، آن جامعه را ازنظر صفت‌متغیّر، «همگن» درنظر می‌گیرند.

مثلاً جامعه‌ای که در مثال بالا آورده شد، نسبت به صفت متغیّر، همگن می‌باشد.

وقتی به‌عنوان اندازهٔ تغییرپذیری، متوسط قدرمطلق انحرافات (A.D)، مورد استفاده قرار می‌گیرد، مشخّصه‌نسبی پراکندگی به‌صورت انحراف‌خطی نسبی، محاسبه می‌شود :

۱ــ Coefficient of variation

انحراف خطی‌نسبی

نسبت متوسّط قدرمطلق انحرافات بر متوسط حسابی صفت متغیّر که برحسب درصد بیان شده باشد، «انحراف‌خطی‌نسبی» نامیده می‌شود. اگر آن‌را با $C.V_{AD}$ نشان دهیم، به‌صورت زیر نوشته می‌شود :

$$C.V_{AD}=\frac{AD}{\overline{X}}\times ۱۰۰ \qquad (۱۵)$$

که در آن :

AD ــ متوسط قدرمطلق انحرافات صفت متغیّر است.

$\overline{X}$ ــ متوسط حسابی صفت متغیّر، می‌باشد.

مثال ۹: انحراف خطی نسبی را برای صفت متغیّر در مثال ۲، محاسبه می‌کنیم که برای آن $\overline{X}=۱۵/۸۵$ و $AD=۲/۱۸$ قبلاً محاسبه شده است.

$$C.V_{AD}=\frac{AD}{\overline{X}}\times ۱۰۰$$

$$=\frac{۲/۱۸}{۱۵/۸۵}\times ۱۰۰=۱۳/۷۵\text{-}$$

درحالاتی که استفاده از مشخّصه‌های تغییرپذیری (پراکندگی) مذکور، امکان‌پذیر نباشد، مناسب است از مشخّصه‌نسبی پراکندگی دیگری به‌نام «انحراف چارکی نسبی» استفاده گردد.

انحراف چارکی نسبی

نسبت انحراف چارکی بر میانهٔ صفت متغیّر «انحراف چارکی نسبی» نامیده می‌شود که آن‌را با q نشان می‌دهیم :

$$q=\frac{QD}{Me}\times ۱۰۰ \qquad (۱۶)$$

مثال ۱۰: انحراف چارکی نسبی را برای صفت متغیّر درمثال ۷ محاسبه می‌کنیم که برای آن : $QD=۸۳۴۱/۵$ و $Me=۵۵۴۸۴$ می‌باشد.

$$q=\frac{QD}{Me}\times ۱۰۰=\frac{۸۳۴۱/۵}{۵۵۴۸۴}\times ۱۰۰=۱۵\text{-}$$

مشخّصه چولگی (انحراف از قرینگی)

برای پی‌بردن به خصوصیّات عمومی توزیع، اغلب، مشخّصه‌های انحراف از قرینگی یا چولگی محاسبه می‌شود.

قبل از اینکه مشخّصه‌های چولگی را معرفی کنیم، باید بدانیم توزیع قرینه چیست، تا براساس آن، غیرقرینه‌بودن، مشخص گردد.

هرگاه فراوانیهای متعلق به دو مقدار صفت که از مرکز توزیع، در فاصله‌های مساوی قرار دارند، یکسان باشند، آن‌توزیع را «قرینه» می‌خوانند. در غیر این صورت، توزیع را «غیرقرینه» یا «چوله» می‌نامند.

مثلاً در توزیع فراوانیهایی که با جدول زیر بیان شده است، فراوانیها، نسبت به وسط توزیع، دوبه‌دو باهم برابرند:

X	۵	۱۰	۱۵	۲۰	۲۵
F_i	۱۰	۴۰	۸۰	۴۰	۱۰

درصورتی که درتوزیع فراوانیهایی که در جدول زیر بیان شده است، فراوانیها برای مقادیر صفت که نسبت به وسط توزیع در فاصله‌های مساوی قرار گرفته‌اند، یکسان نیستند.

X_i	۲	۴	۶	۸	۱۰	۱۲
F_i	۱۰	۲۰	۴۰	۴۵	۲۵	۱۰

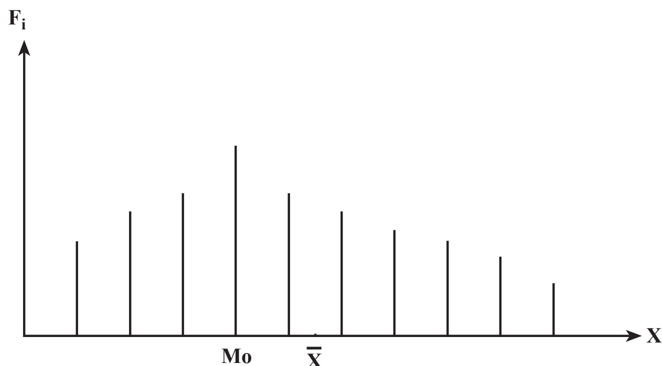
برای مقایسه درجه‌دوری توزیعها ازحالت قرینه‌بودن، از مشخّصه‌نسبی که «ضریب چولگی» نام دارد، استفاده می‌شود.

ــ ضریب چولگی پیرسن: در فصل گذشته، ملاحظه شد که برای توزیعهای قرینه، مقدار میانگین، میانه و نما باهم برابرند. درارتباط با آن، ساده‌ترین مشخّصه انحراف از قرینگی یا چولگی که ضریب چولگی پیرسن نام دارد، بر مناسبات بین مشخّصه‌های مرکزی استوار گردیده است.

هرچه تفاوت بین مشخّصه‌های مرکزی $(\overline{X} - Mo)$، بیشتر باشد، چولگی نیز بیشتر است. ضریب چولگی پیرسن به صورت زیر تعریف می‌شود:

$$P_{AS} = \frac{\overline{X} - Mo}{.} \qquad (۱۷)$$

کمیّت P_{AS} ، می‌تواند مثبت، منفی یا صفر باشد. مقدار مثبت این ضریب، وجود چولگی راست را نشان می‌دهد یعنی دامنهٔ راست توزیع نسبت به نما، بیشتر کشیده است تا دامنهٔ چپ آن (مانند نمودار شکل ۱) :

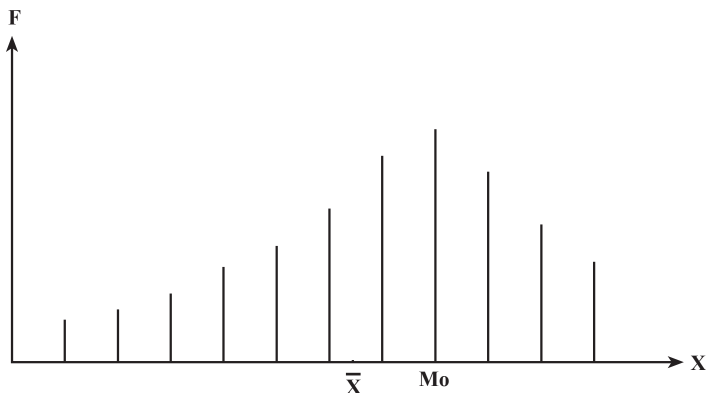


شکل ۱ــ وجود چولگی راست

در چولگی راست، بین مشخّصه‌های مرکزی رابطهٔ زیر برقرار است :

$$Mo < Me < \overline{X} \quad (۱۸)$$

علامت منفی در این ضریب، گواه وجود چولگی چپ است که در شکل ۲ نشان داده شده است :



شکل ۲ــ وجود چولگی چپ

در این حالت، بین مشخّصه‌های مرکزی، رابطهٔ زیر برقرار است :

Mo # Me # $\overline{X}$	(۱۹)

معمولاً وقتی ضریب چولگی کوچکتر از ۰/۱ باشد، توزیع را تقریباً قرینه و وقتی که ضریب چولگی بزرگتر از ۰/۵ باشد آن‌را خیلی چوله درنظر می‌گیرند. طبیعی است که در توزیعهای قرینه، ضریب چولگی مساوی با صفر خواهد بود.

مثال ۱۱: ضریب چولگی پیرسن (P_{AS}) را برای توزیع زیر محاسبه می‌کنیم :

X	۱۰–۱۵	۱۵–۲۰	۲۰–۲۵	۲۵–۳۰	۳۰–۳۵	–
F_i	۴	۱۰	۲۰	۵	۱	۴۰

برای محاسبهٔ ضریب چولگی پیرسن، میانگین حسابی، نما و انحراف معیار توزیع را باید به‌دست آورد. تمامی محاسبات در جدول ۷ آورده شده است.

جدول ۷

X	F_i	X_i	F_iX_i	$F_iX_i^۲$
۱۰–۱۵	۴	۱۲/۵	۵۰	۶۲۵
۱۵–۲۰	۱۰	۱۷/۵	۱۷۵	۳۰۶۲/۵
۲۰–۲۵	۲۰	۲۲/۵	۴۵۰	۱۰۱۲۵
۲۵–۳۰	۵	۲۷/۵	۱۳۷/۵	۳۷۸۱/۲۵
۳۰–۳۵	۱	۳۲/۵	۳۲/۵	۱۰۵۶/۲۵
–	۴۰	–	۸۴۵	۱۸۶۵۰

$$\overline{X} = \frac{.\ F_iX_i}{N} = \frac{۸۴۵}{۴۰} = ۲۱/۱۲۵$$

$$V(X) = \frac{.\ F_iX_i^{۲} - \frac{(.\ F_iX_i)^{۲}}{N}}{N}$$

$$= \frac{۱۸۶۵۰ - \frac{(۸۴۵)^{۲}}{۴۰}}{۴۰} = \frac{۷۹۹/۳۷۵}{۴۰} = ۱۹/۹۸$$

$$= .+ \quad \sqrt{V(X)} = \sqrt{۱۹/۹۸} = ۴/۴۷$$

$$Mo = X_i + \frac{F_i - F_{i-۱}}{(F_i - F_{i-۱}) + (F_i - F_{i+۱})} \times I$$

$$= ۲۰ + \frac{۲۰-۱۰}{(۲۰-۱۰)+(۲۰-۵)} \times ۵$$

$$= ۲۰ + \frac{۵۰}{۲۵} = ۲۲$$

درنتیجه، ضریب چولگی پیرسن طبق فرمول۱۷، به‌صورت زیر محاسبه می‌شود :

$$P_{AS} = \frac{\overline{X} - Mo}{.} = \frac{۲۱/۱۲۵ - ۲۲}{۴/۴۷} = -۰/۱۹$$

ولی باید متذکر شد که ضریب چولگی پیرسن همیشه نمی‌تواند مورد استفاده قرار گیرد. مثلاً اگر توزیع دو نما داشته باشد، این ضریب نمی‌تواند محاسبه گردد.

ــ ضریب چولگی بر پایه میانگین توان سوم انحرافات:

یکی از متداول‌ترین ضرایب چولگی که نقص ضریب چولگی پیرسن را ندارد و از تمامی عناصر توزیع (یعنی از مقادیر صفت و فراوانیها) در محاسبه آن استفاده می‌شود ضریب چولگی بر پایه میانگین توان سوم انحرافات می‌باشد که با M_{AS} نشان داده می‌شود، به صورت زیر بیان می‌شود :

$$M_{AS} = \frac{\frac{.\ F_i(X_i - \overline{X})^۳}{N}}{.^{۳}} = \frac{.\ F_i(X_i - \overline{X})^۳}{N.^{۳}} \qquad (۲۰)$$

مثال ۱۲: ضریب چولگی بر پایهٔ میانگین توان سوم انحرافات را برای توزیع فراوانیهای صفت متغیّر که با جدول زیر بیان شده است، محاسبه می‌کنیم :

X	۳	۴	۵	۶	۷	۸	−
F_i	۱	۶	۱۳	۱۱	۱۰	۹	۵۰

تمامی محاسبات در جدول ۸، آورده شده است :

جدول ۸

X	F_i	F_iX_i	$F_iX_i^۲$	$X_i - \overline{X}$	$(X_i - \overline{X})^۳$	$F_i(X_i - \overline{X})^۳$
۳	۱	۳	۹	−۳	−۲۷	−۲۷
۴	۶	۲۴	۹۶	−۲	−۸	−۴۸
۵	۱۳	۶۵	۳۲۵	−۱	−۱	−۱۳
۶	۱۱	۶۶	۳۹۶	۰	۰	۰
۷	۱۰	۷۰	۴۹۰	۱	۱	۱۰
۸	۹	۷۲	۵۷۶	۲	۸	۷۲
−	۵۰	۳۰۰	۱۸۹۲	−	−	−۶

$$\overline{X} = \frac{۳۰۰}{۵۰} = ۶$$

$$V(X) = \frac{.\ F_iX_i^۲}{N} - \overline{X}^۲$$

$$= \frac{۱۸۹۲}{۵۰} - ۶^۲ = ۳۷/۸۴ - ۳۶ = ۱/۸۴$$

$$= .+\ \sqrt{V(X)} = \sqrt{۱/۸۴} = ۱/۳۶$$

$$.\ ^۳ = ۱/۳۶ \times ۱/۸۴ = ۲/۵۰$$

با استفاده از کمیّتهای محاسبه‌شده، طبق فرمول ۲۰، ضریب‌چولگی را محاسبه می‌کنیم :

$$M_{AS} = \frac{\frac{.\ F_i(X_i - \overline{X})^۳}{N}}{.\ ^۳} = \frac{\frac{-۶}{۵۰}}{۲/۵} = -۰/۰۴۸$$

بنابراین، چون ضریب چولگی، کوچکتر از ۰/۱ به‌دست آمده است، می‌توان گفت، توزیع تقریباً قرینه است.

سؤالها و تمرینها

۱ــ چرا در مطالعهٔ صفت‌متغیّر، برای اعضای جامعه، تنها به کمیّتهای متوسط نمی‌توان اکتفا کرد؟

۲ــ اندازهٔ پراکندگی چیست؟ ساختن اندازهٔ پراکندگی انحرافات مقادیر صفت از متوسط صفت، چه اهمیّتی دارد؟

۳ــ طول دامنهٔ تغییرات چیست؟

۴ــ طول دامنهٔ تغییرات را برای داده‌های زیر به‌دست آورید :

۱۵،۱۰،۱۳،۶،۱۱،۷،۵،۹،۱۰، ۸،۷،۴،۳

۵ــ طول دامنه تغییرات را برای طول‌قد دانش‌آموزان که در جدول زیر داده شده است، محاسبه کنید.

طول قد X	۱۴۴–۱۴۸	۱۴۸–۱۵۲	۱۵۲–۱۵۶	۱۵۶–۱۶۰	۱۶۰–۱۶۴	
فراوانی F_i	۲	۶	۱۰	۱۵	۷	۴۰

۶ــ نقص طول دامنهٔ تغییرات به‌عنوان یک مشخّصه پراکندگی در چیست؟

۷ــ چرا از انحرافات مقادیر صفت، نسبت به متوسط حسابی، نمی‌توانیم متوسط انحرافات را محاسبه کنیم؟

۸ــ متوسط قدرمطلق انحرافات را تعریف کنید.

۹ــ متوسط قدرمطلق انحرافات را برای توزیع قد دانش‌آموزان در تمرین ۵ محاسبه کنید.

۱۰ــ چرا در عمل، از متوسّط قدرمطلق انحرافات، کمتر استفاده می‌شود؟

۱۱ــ متوسط مجذور انحرافات یا واریانس، چه مشخّصه‌ای است؟

۱۲ــ واریانس به‌عنوان مشخّصهٔ پراکندگی، برای برطرف کردنِ چه مشکلی در آمار، وارد گردیده است؟

۱۳ــ فرمولهای کاربردی یا فرمولهای محاسباتی واریانس، کدامها هستند؟ آنها را بنویسید.

۱۴ــ واریانس را برای طول قد دانش‌آموزان درتمرین ۵ محاسبه کنید.

۱۵ــ واریانس را برای اندازه‌های صفت که در زیر داده شده‌اند، محاسبه کنید.

X:۳,۵,۳, ۶,۷,۲,۸,۱۰

۱۶ــ چرا به‌جای واریانس، برای نشان‌دادن پراکندگی صفت، از انحراف معیار، استفاده می‌شود؟

۱۷ــ انحراف معیار چیست؟ تعریف کنید.

۱۸ــ انحراف معیار را برای تمرین ۵ (طول قد دانش‌آموزان)، محاسبه کنید.

۱۹ــ برای توزیع فراوانیهای صفت که در زیر داده شده، انحراف معیار و متوسط قدرمطلق انحرافات را محاسبه کرده، آنها را باهم مقایسه کنید. به‌نظر شما کدام دقیق‌تر است؟

X	۲	۴	۶	۸	۱۰	۱۲	ــ
F_i	۴	۷	۱۰	۶	۲	۱	۳۰

۲۰ــ انحراف چارکی چیست؟ تعریف کنید.

۲۱ــ لزوم واردکردن انحراف چارکی به‌عنوان مشخّصه پراکندگی، در چیست؟

۲۲ــ چارکها، چه مفهومی را بیان می‌کنند؟

۲۳ــ چارک اول چیست؟

۲۴ــ چارک دوم چیست؟

۲۵ــ چارک سوم چیست؟

۲۶ــ انحراف چارکی را برای توزیع فراوانیهای زیر محاسبه کنید :

X	۴ــ۸	۸ــ۱۲	۱۲ــ۱۶	۱۶ــ۲۰	۲۰ــ۲۴	۲۴ به بالا
F_i	۳	۵	۸	۱۲	۶	۶

۲۷ــ مشخّصه‌های نسبی پراکندگی، چه مفهومی دارند؟

۲۸ــ چرا برای بیان پراکندگی صفت، از مشخّصه‌های نسبی پراکندگی استفاده می‌شود؟

۲۹ــ ضریب تغییرات چیست؟

۳۰ــ ضریب تغییرات را برای تمرین ۱۵ محاسبه کنید.

۳۱ــ ضریب تغییرات را برای تمرین ۱۹ محاسبه کنید.

۳۲ــ ضریب «انحراف خطی نسبی»، چیست؟ در چه مواردی از آن استفاده می‌کنند؟
۳۳ــ انحراف خطی نسبی را برای تمرین ۵ محاسبه کنید.
۳۴ــ انحراف خطی نسبی را برای تمرین ۱۹ محاسبه کنید.
۳۵ــ انحراف چارکی نسبی چیست؟ در چه مواردی از آن استفاده می‌کنند؟
۳۶ــ انحراف چارکی نسبی را برای تمرین ۲۶ محاسبه کنید.
۳۷ــ برای پی‌بردن به خصوصیّات عمومی توزیعها، از چه نوع مشخّصه‌هایی استفاده می‌شود؟
۳۸ــ چه نوع توزیعی را، توزیع قرینه گویند؟ تعریف کنید.
۳۹ــ توزیع غیرقرینه یا توزیع چوله، چه نوع توزیعی است؟
۴۰ــ مشخّصه چولگی یا ضریب چولگی چیست؟ چه خصوصیّاتی از توزیع صفت را بیان می‌کند؟
۴۱ــ ضریب چولگی پیرسن، بر چه اساسی ساخته می‌شود؟
۴۲ــ ضریب چولگی پیرسن را برای تمرین ۱۵محاسبه کنید.
۴۳ــ ضریب چولگی پیرسن را برای تمرین ۱۹ محاسبه کنید.
۴۴ــ ضریب چولگی منفی، چگونه تعبیر می‌شود؟
۴۵ــ ضریب چولگی مثبت، چگونه تعبیر می‌شود؟
۴۶ــ ضریب چولگی براساس متوسط توان سوم انحرافات چیست؟ فرمول آن‌را بنویسید.
۴۷ــ ضریب چولگی براساس متوسط توان سوم انحرافات را برای تمرین ۱۵ محاسبه کنید.
۴۸ــ ضریب چولگی براساس متوسّط توان سوم انحرافات را برای تمرین ۱۹ محاسبه کنید.

خودآزمونهای چهارگزینه‌ای ✔



۱ــ کدام‌یک از مشخّصه‌های پراکندگی زیر، ازنظر دقت، اعتبار بیشتری دارد؟

الف ــ دامنهٔ تغییرات　　　ب ــ واریانس

ج ــ متوسط قدرمطلق انحرافات　　　د ــ انحراف معیار

۲ــ از روی یک نمونه $n=۱۰۰$ در جامعه، کمیّتهای زیر محاسبه شده است :

$$n=۱۰۰ \quad . \quad F_iX_i=۱۲۰۰ \quad . \quad F_i(X_i-\overline{X})^۲=۳۶$$

واریانس جامعه کدام است؟

الف ــ ۰/۳۶　　ب ــ ۰/۱۶　　ج ــ ۰/۶　　د ــ ۰/۱۲

۳ــ اگر مجموع توان دوم، ۱۰ اندازه از صفت متغیّر ۱۷۰۰ = X^2 . و میانگین آنها ۱۲ = $\bar{X}$ باشد، واریانس X کدام است؟

الف ــ ۶　　ب ــ ۱۶　　ج ــ ۲۶　　د ــ صفر

۴ــ با معلوم بودن ۱۰ = $\bar{X}$ و ۱۲۰۰ = X^2 . ، ۲۰ = V(X) ، حجم جامعه چقدر بوده است؟

الف ــ ۸　　ب ــ ۲۰　　ج ــ ۱۲　　د ــ ۱۰

۵ــ انحراف چارکی برای توصیف کدام مشخّصهٔ صفت، به‌کار می‌رود؟

الف ــ مشخّصهٔ مرکزی　　ب ــ مشخّصهٔ پراکندگی

ج ــ مشخّصهٔ چولگی　　د ــ هیچ‌کدام

۶ــ اگر بین مشخّصه‌های مرکزی رابطهٔ $\bar{X}$ # Me # Mo برقرار باشد، کدام گزینه برای صفت متغیّر درست است؟

الف ــ چولگی به چپ است.　　ب ــ چولگی ندارد.

ج ــ چولگی به راست است.　　د ــ توزیع قرینه است.

۷ــ کدام‌یک از روابط زیر، فرمول ضریب چولگی بر پایه میانگین توان سوم انحرافات است؟

الف ــ $\dfrac{.\ F_i(X_i-\bar{X})^3}{.^3}$　　ب ــ $\dfrac{\dfrac{.\ F_i(X_i-\bar{X})^2}{N}}{.^3}$

ج ــ $\dfrac{.\ F_i(X_i-\bar{X})^3}{N}$　　د ــ $\dfrac{\dfrac{.\ F_i(X_i-\bar{X})^3}{N}}{.^3}$

۸ــ کدام گزینه تعریف ضریب تغییرات صفت می‌باشد؟

الف ــ نسبت میانگین بر انحراف معیار صفت

ب ــ نسبت میانه بر انحراف معیار صفت

ج ــ نسبت انحراف معیار بر میانه صفت

د ــ نسبت انحراف معیار بر میانگین صفت

۹ــ انحراف خطی نسبی از کدام گزینه به‌دست می‌آید؟

الف ــ $\frac{AD}{\overline{X}}$ ب ــ $\frac{QD}{\overline{X}}$ ج ــ $\frac{\cdot}{Me}$ د ــ $\frac{V(X)}{\overline{X}}$

۱۰ــ کدام یک از گزینه‌های زیر ضریب انحراف چارکی را بیان می‌کند؟

الف ــ $\frac{Q_3 - Q_2}{2}$ ب ــ $\frac{QD}{Me}$

ج ــ $\frac{Q_3 - Q_1}{2}$ د ــ $\frac{Q_3 - Q_1}{Me}$

فصل پنجم

توزیع صفت متغیّر کیفی

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند :

۱ــ جامعه را برحسب صفات کیفی گروه‌بندی کند.

۲ــ مشخّصه‌های مرکزی را برای صفات کیفی محاسبه نماید.

۳ــ مشخّصه‌های پراکندگی را برای صفات کیفی محاسبه نماید.

۴ــ جامعه‌ها را با استفاده از مشخّصه‌های پراکندگی باهم مقایسه کند و همگنی آنها را بیان نماید.

۵ــ جامعه‌ها را از نظر دو صفت متغیّر کیفی که هر یک می‌توانند حالات مختلفی را اختیار کنند، گروه‌بندی نماید.

۶ــ جدول توزیع فراوانیهای نسبی را برحسب حالات دو صفت کیفی، تشکیل دهد.

توزیع صفت کیفی

همان‌طور که قبلاً در طبقه‌بندی صفتهای متغیّر گفته شد، صفت متغیّر کیفی، صفتی است که آن‌را برای اعضای جامعه با عدد یا کمیت نتوان بیان نمود. برای بیان توزیع صفت کیفی فقط می‌توان حالتهای آن‌را از هم متمایز ساخته، تعداد اعضای هر یک از حالتها را با فراوانیهای (مطلق یا نسبی) بیان نمود که جدول توزیع فراوانی آن به‌صورت زیر نشان داده می‌شود :

جدول ۱

حالات صفت **X**	$A_1 \quad A_2 \quad A_3 \quad \dots \quad A_s$	–
فراوانی $\mathbf{F_i}$	$F_1 \quad F_2 \quad F_3 \quad \dots \quad F_s$	$= N$

A_1 ، A_1 ، ...، A_S حالتهای صفت کیفی برای اعضای جامعه می‌باشد. مثلاً اگر بخواهیم

طبقه‌بندی کارکنان یک مؤسسه را برحسب مشاغل بیان کنیم، توزیع فراوانیهای کارکنان برحسب صفت شغل به‌صورت جدول زیر بیان می‌شود :

جدول ۲

مشاغل X	مستخدم	نگهبان	انباردار	راننده	کارمند	مدیر عامل ...
فراوانی F_i	۳	۱	۲	۳	۶	$N = ۱...$

و یا اگر بخواهیم طبقه‌بندی بیماران یک بیمارستان را برحسب نوع بیماری نشان دهیم، از جدول توزیع فراوانی بیماران برحسب نوع بیماری (X) استفاده می‌کنیم :

جدول ۳

نوع بیماری X	بیماران عفونی	بیماران قلبی	بیماران عصبی	بیماران گوارشی	...	–
فراوانی F_i	۱۰	۱۵	۶	۱۲	...	$=N$

بر طبق معمول در مطالعه صفت کیفی، آن حالت از صفت متغیّر کیفی که موردنظر می‌باشد از سایر حالتهای صفت کیفی مجزا کرده، به‌صورت توزیع دو حالتی بیان می‌کنند. در این صورت، توزیع صفت کیفی را با دو حالت متضاد می‌نامند حالت موردنظر را با A و بقیه حالتها را با $\overline{A}$ نشان داده، توزیع فراوانیها را برای صفت دو حالت متضاد به‌صورت زیر بیان می‌کنند :

جدول ۴

X	$\overline{A}$	A	–
$\mathbf{F_i}$	$F_{\overline{A}}$	F_A	$=N$

یا به‌صورت توزیع فراوانیهای نسبی :

جدول ۵

X	$\overline{A}$	A	–
$\mathbf{f_i}$	$f_{\overline{A}}$	f_A	=۱

مثال ۱ : در مجموعهٔ ۴۰ سند صادره در یک واحد حسابداری، برای تعیین کیفیت اسناد صادره، اسناد تحت بازرسی قرارگرفته‌اند و معلوم شده است که **۲۹** فقره از آنها «بدون نقص» (حالت $\overline{A}$) و بقیه «نقص‌دار» (حالت A) بوده‌اند. بنابراین مجموعه اسناد را می‌توان برحسب نقص‌دار و

بدون نقص به دو گروه تقسیم کرد که نتیجه آن در جدول۶ بیان شده است :

جدول ۶

حالات صفت	بدون نقص	نقص‌دار	
	$\overline{A}$	A	—
فراوانی	۲۹	۱۱	۴۰ =

اگر بخواهیم جدول توزیع فراوانیهای نسبی را برای مجموعه اسناد بیان کنیم، کافی است هر یک از فراوانیهای مطلق را به N یعنی ۴۰ تقسیم کنیم، در نتیجه خواهیم داشت :

حالات صفت	$\overline{A}$	A	—
فراوانی نسبی f_i	$\frac{۲۹}{۴۰}$	$\frac{۱۱}{۴۰}$	۱=

مثال ۲: فرض کنیم جامعهٔ مورد مطالعه $N = ۲۰۰$ واحد محصول تولیدشده در یک مؤسّسه تولیدی باشد که واحدهای محصول از سوی کارشناسان، ازنظر کیفیت تحت بازرسی قرار گرفته و براساس آن، محصول به چهار طبقه، درجه‌بندی شده است :

پس از شمارش واحدهای محصول در هرگروه، نتایج گروه‌بندی به‌صورت جدول ۷ درآمده است.

جدول ۷

کیفیت‌محصول	**درجه ۱**	**درجه ۲**	**درجه۳**	**درجه ۴**	—
فراوانی	۶۰	۸۰	۴۰	۲۰	۲۰۰

اگر فراوانیهای هرگروه را بر حجم جامعه ($N = ۲۰۰$) تقسیم نماییم، جدول بالا به‌صورت زیر درمی‌آید (جدول۸) :

جدول ۸

کیفیت محصول	**درجه ۱**	**درجه ۲**	**درجه ۳**	**درجه ۴**	—
فراوانی نسبی	۰/۳۰	۰/۴۰	۰/۲۰	۰/۱۰	۱

این دو جدول، توزیع فراوانیهای واحدهای محصول را برحسب کیفیّت بیان می کند.

محاسبهٔ مشخّصه‌های عددی برای صفات کیفی

از آنجا که صفت متغیّر، کیفی است و با حالت بیان می‌شود، نمی‌توان برای آنها مشخّصه‌های عددی محاسبه نمود. برای این‌کار لازم است، حالات صفات کیفی را به صورت کمّی بیان کرد و با آنها مانند صفات کمّی برخورد نمود.

فرض کنیم صفت متغیّر کیفی، حالتهای A_1 ، A_2 ، ...، A_S را بتواند در جامعه اختیار کند. اگر حالتهای مختلف صفت را به دو گروه، حالت موردنظر (A) و حالت غیر موردنظر $(\overline{A})$ تقسیم‌بندی کنیم، درنتیجه، توزیع صفت به صورت زیر درخواهد آمد :

$$X\begin{pmatrix} A & \overline{A} \\ F_A & F_{\overline{A}} \end{pmatrix}$$

حال با تبدیل حالت موردنظر (A) به عدد ۱ و حالت غیر موردنظر $(\overline{A})$ به عدد صفر، توزیع به‌صورت زیر درخواهد آمد :

$$X\begin{pmatrix} ۱ & ۰ \\ F_A & F_{\overline{A}} \end{pmatrix}$$

اینک می‌توان برای چنین صفتی که به‌صورت کمّی درآمده است، مشخّصه‌مرکزی و مشخّصه پراکندگی محاسبه نمود.

میانگین صفت کیفی خواهد شد :

$$\overline{X} = \frac{F_A}{N} = f_A \qquad (۳)$$

بنابراین، میانگین صفت‌کیفی، فراوانی نسبی حالت موردنظر خواهد بود.

واریانس صفت‌کیفی، به‌صورت زیر به‌دست می‌آید :

$$V(X) = f_A . f_{\overline{A}} \qquad (۴)$$

بنابراین، واریانس صفت کیفی، حاصل‌ضرب فراوانی نسبی حالت موردنظر و فراوانی نسبی حالت غیرموردنظر خواهد بود.

مثال ۳: فرض کنید در مثال ۲ می‌خواهیم متوسط محصولات درجهٔ ۱ (A) را محاسبه کنیم.

برای این‌کار، کافی‌است محصولات درجهٔ ۱ را، حالت موردنظر (A)، درنظر گرفته و بقیّه محصولات را، حالت غیرموردنظر ($\overline{A}$)، درنظر بگیریم.

درنتیجه، توزیع به‌صورت زیر درمی‌آید :

	محصولات درجه ۱	سایر محصولات	
X	A	$\overline{A}$	
	۶۰	۱۴۰	۲۰۰
X	۱	۰	
	۶۰	۱۴۰	۲۰۰

بنابراین، متوسط محصولات درجهٔ ۱، خواهد شد :

$$\overline{X} = \frac{۶۰}{۲۰۰} = ۰/۳۰$$

برای محاسبهٔ واریانس، از فرمول ۴ استفاده می‌کنیم :

$$V(X) = f_A . f_{\overline{A}}$$
$$= ۰/۳۰ \times ۰/۷ = ۰/۲۱$$

کمترین مقدار واریانس صفت کیفی، صفر و بیشترین مقدار آن ۰/۲۵ خواهد بود. غالباً برای پی‌بردن به همگنی جامعه‌های با صفات کیفی، از واریانس آنها استفاده می‌شود. اگر مقدار واریانس صفت کیفی، برابر صفر باشد، گویند جامعه از همگنی کامل برخوردار است و اگر مقدار واریانس صفت کیفی برابر ۰/۲۵ باشد، گویند جامعه از ناهمگنی کامل برخوردار است. بنابراین هرچه مقدار واریانس صفت کیفی به صفر نزدیکتر باشد، نشانهٔ همگنی بیشتر در جامعه است و هرچه مقدار واریانس صفت‌کیفی به ۰/۲۵ نزدیکتر باشد، نشانهٔ ناهمگنی بیشتر در جامعه خواهد بود.

معمولاً در مقایسهٔ همگنی جامعه‌ها، با استفاده از واریانس آنها، می‌توان نسبت به همگنی آنها قضاوت نمود.

مثال ۴: دو گروه محصول یکسان را برحسب مرغوبیّت، دسته‌بندی نموده‌ایم و نتایج آن، به‌صورت زیر به‌دست آمده است :

I

کیفیت محصول	مرغوب $\overline{A}$	نامرغوب A
f_i فراوانی‌نسبی	۰/۷۵	۰/۲۵

II

کیفیت محصول	مرغوب $\overline{A}$	نامرغوب A
f_i فراوانی‌نسبی	۰/۸۲	۰/۱۸

تعیین کنید، کدام گروه، همگن‌تر است؟

برای پاسخ این سؤال، لازم است واریانس هریک از گروه‌ها را محاسبه کنیم، هرکدام واریانس کمتری نسبت به دیگری داشته باشد، آن گروه محصول همگنی بیشتری دارد.

$$V(X_۱) = f_A . f_{\overline{A}} = ۰/۷۵ \times ۰/۲۵ = ۰/۱۸۷۵$$

$$V(X_۲) = f_A . f_{\overline{A}} = ۰/۸۲ \times ۰/۱۸ = ۰/۱۴۷۶$$

ملاحظه می‌شود که واریانس گروه دوم کمتر از واریانس گروه اول شده، بنابراین نتیجه می گیریم که گروه دوم ازنظر کیفیّت همگن‌تر است.

گاهی اوقات، توزیع صفت کیفی برحسب فراوانیهای مطلق داده می‌شود. برای محاسبهٔ واریانس، لازم است، ابتدا، فراوانیهای نسبی را برای حالتهای صفت محاسبه نموده، سپس با استفاده از فراوانیهای نسبی، واریانس را محاسبه کنیم.

مثال ۵: جمعیّت دو روستا برحسب وضع سواد، به‌صورت جداول زیر گروه‌بندی شده‌اند :

روستای A

وضع سواد	باسواد	بی‌سواد	جمع
فراوانی	۲۴۹	۵۱	۳۰۰

روستای B

وضع سواد	باسواد	بی‌سواد	جمع
فراوانی	۳۲۴	۷۶	۴۰۰

کدام‌یک از دو روستای فوق از نظر سواد همگن‌تر هستند؟

برای تعیین چگونگی همگنی دو روستای فوق ازنظر وضع سواد، می‌باید واریانس هریک از آنها را محاسبه کرد. واریانس صفت کیفی، فقط با فراوانیهای نسبی می‌تواند محاسبه شود، از این‌رو، ابتدا فراوانیهای نسبی را برای هر دو روستا محاسبه نموده، سپس با استفاده از آنها، واریانس را محاسبه می‌کنیم.

روستای B

وضع سواد	باسواد	بی‌سواد	جمع
F_i	۳۲۴	۷۶	۴۰۰
f_i	۰/۸۱	۰/۱۹	۱

روستای A

وضع سواد	باسواد	بی‌سواد	جمع
F_i	۲۴۹	۵۱	۳۰۰
f_i	۰/۸۳	۰/۱۷	۱

$V(X)_B = ۰/۸۱ \times ۰/۱۹ = ۰/۱۵۳۹$ $\quad$ $V(X)_A = ۰/۸۳ \times ۰/۱۷ = ۰/۱۴۱۱$

از این‌رو، می‌توان ادّعا نمود که روستای A ازنظر وضع سواد، همگن‌تر از روستای B است، چون واریانس کمتری از روستای B دارد.

مطالعه جامعه ازنظر دو صفت کیفی A و B

حال، جامعه‌ای را درنظر می‌گیریم که عناصر آن ازنظر دو صفت A و B که هر یک با حالات متضاد مشاهده شده‌اند، مورد مطالعه قرار گرفته‌اند.

در این وضعیّت ممکن است یکی از چهارحالت زیر وجود داشته باشد :

ــ برای عنصر جامعه، هم صفت A و هم صفت B مشاهده شده است. که این حالت ترکیبی را، برای عنصر مشاهده شده با علامت AB نشان می‌دهیم.

ــ برای عنصر جامعه، صفت A مشاهده نشده ولی صفت B مشاهده شده است. این حالت ترکیبی را با نماد $\overline{A}B$ نشان می‌دهیم.

ــ برای عنصر جامعه صفت A مشاهده شده ولی صفت B مشاهده نشده است. این حالت ترکیبی را با نماد $A\overline{B}$ نشان می‌دهیم.

ــ برای عنصر جامعه، صفت A و صفت B مشاهده نشده است، که این حالت ترکیبی را با نماد $\overline{A}\,\overline{B}$ نشان می‌دهیم.

گروه‌بندیی را که نسبت به دو صفت باهم انجام می‌گیرد، «گروه‌بندی مرتبهٔ دوم» خواهیم نامید. گروه‌بندی جامعه برحسب یک صفت متغیّر را «گروه‌بندی مرتبه اول» می‌نامند. حال اگر تعداد عناصر (فراوانیهای) چهار گروه نامبرده را با F_{AB} ، $F_{\overline{A}B}$ ، $F_{A\overline{B}}$ و $F_{\overline{A}\overline{B}}$ نشان دهیم، نتیجه این گروه‌بندی را می‌توان به‌صورت جدول چهارخانه‌ای زیر بیان کرد :

جدول ۹

A \ B	B	$\overline{B}$	جمع
A	F_{AB}	$F_{A\overline{B}}$	F_A
$\overline{A}$	$F_{\overline{A}B}$	$F_{\overline{A}\overline{B}}$	$F_{\overline{A}}$
	F_B	$F_{\overline{B}}$	N

تمامی فراوانیها را در جدول ۹ برحجم جامعه (N) تقسیم کرده، جدول چهارخانه‌ای مشابه را با فراوانیهای نسبی به‌دست می‌آوریم (جدول ۱۰) :

جدول ۱۰

A \ B	B	$\overline{B}$	جمع
A	f_{AB}	$f_{A\overline{B}}$	f_A
$\overline{A}$	$f_{\overline{A}B}$	$f_{\overline{A}\overline{B}}$	$f_{\overline{A}}$
	f_B	$f_{\overline{B}}$	۱

مثال ۶: فرض کنیم جامعهٔ مورد مطالعه، ۱۰۰ کارگر یک مؤسّسه تولیدی است که ازنظر داشتن سواد (صفت A) و داشتن بچه (صفت B) مورد مشاهده قرار گرفته‌اند. پس‌از تعیین داشتن سواد و داشتن بچه برای هر کارگر، نتایج مشاهدات به ۴ گروه طبقه‌بندی گردید :

گروه اول — باسواد و با بچه — AB

گروه دوم — باسواد و بدون بچه — $A\overline{B}$

گروه سوم — بی‌سواد و با بچه — $\overline{A}B$

گروه چهارم — بی‌سواد و بدون بچه — $\overline{A}\overline{B}$

نتایج مشاهدات به‌صورت جدول چهارخانه‌ای زیر بیان می‌شود (جدول ۱۱) :

جدول ۱۱

	بچه‌دار B	بدون بچه $\overline{B}$	جمع
باسواد A	۳۵	۲۵	۶۰
بی‌سواد $\overline{A}$	۳۰	۱۰	۴۰
جمع	۶۵	۳۵	۱۰۰

جدول بالا، توزیع فراوانیهای کارگران را بر حسب دو صفت A و B بیان می‌کند. اگر فراوانیهای مطلق هر گروه را بر حجم جامعه (N=۱۰۰) تقسیم نماییم، جدول توزیع فراوانیهای نسبی کارگران برحسب دو صفت A و B به‌دست می‌آید که تشکیل آن‌را به‌عنوان تمرین، به دانش‌آموزان واگذار می‌کنیم.

به همین شیوه که با دو صفت با حالات متضاد، گروه‌بندی مرتبهٔ دوم انجام می‌گرفت، می‌توان گروه‌بندیهای مراتب بالاتر را نیز برحسب ۳ و بیش از ۳ صفت تشکیل داد.

در حالت دو صفت A و B که هریک از آنها با بیش از دو حالت ($A_۱$ ، $A_۲$ ، $A_۳$ ، ....، A_S) برای صفت A و ($B_۱$ ، $B_۲$ ، $B_۳$ ، ....، B_t) برای صفت B مشاهده می‌شوند، درنتیجه گروه‌بندی مرتبه اول، S گروه برای صفت A و t گروه برای صفت B، تشکیل خواهد شد. نتایج این گروه‌بندیها با جدول دوبعدی (یا دوطرفه)، که هریک از بُعدها را یکی از صفتهای متغیّر کیفی، تشکیل می‌دهند، بیان خواهد شد (جدول ۱۲) :

جدول ۱۲

A \ B	$B_۱$	$B_۲$	$B_۳$	⋯	B_t	جمع
$A_۱$	$F_{۱۱}$	$F_{۱۲}$	$F_{۱۳}$	⋯	$F_{۱t}$	$F_{A_۱}$
$A_۲$	$F_{۲۱}$	$F_{۲۲}$	$F_{۲۳}$	⋯	$F_{۲t}$	$F_{A_۲}$
⋮	⋮	⋮	⋮		⋮	⋮
A_S	$F_{S۱}$	$F_{S۲}$	$F_{S۳}$	⋯	F_{St}	F_{AS}
جمع	$F_{B_۱}$	$F_{B_۲}$	$F_{B_۳}$	⋯	F_{Bt}	N

این جدول، صورت تعمیم‌یافتهٔ جدول چهارخانه‌ای ۹ می‌باشد و نام «جدول توافق» به خود گرفته است.

مثال ۷: جامعهٔ مورد مطالعه کارگران یک مؤسّسهٔ تولیدی با حجم $N = ۵۰$ هستند که برحسب درجهٔ مهارت (A) و میزان تحصیلات (B) مورد مشاهده قرار می‌گیرند.

درجه مهارت (A) می‌تواند به ۴ حالت :

$A_۱$ ــ کارگر ساده

$A_۲$ ــ کارگر نیمه‌ماهر

$A_۳$ ــ کارگر ماهر

$A_۴$ ــ استادکار

و میزان تحصیلات می‌تواند به ۴ حالت :

$B_۱$ ــ بی‌سواد

$B_۲$ ــ تحصیلات ابتدایی

$B_۳$ ــ تحصیلات متوسطه

$B_۴$ ــ تحصیلات دیپلم

درنظر گرفته شود. عناصر جامعه (کارگران) برحسب دو صفت متغیّر A و B مورد مشاهده قرار گرفته‌اند و نتایج به‌صورت جدول زیر طبق حالتهای فوق گروه‌بندی گردیده است :

جدول ۱۳

A \ B	بی‌سواد	ابتدایی	متوسطه	دیپلم	جمع
کارگر ساده	۶	۵	۱	۰	۱۲
کارگر نیمه‌ماهر	۳	۵	۳	۱	۱۲
کارگر ماهر	۲	۴	۷	۳	۱۶
استادکار	۰	۲	۵	۳	۱۰
جمع	۱۱	۱۶	۱۶	۷	۵۰

جدول بالا، توزیع فراوانیهای کارگران مؤسّسه مذکور را برحسب دو صفت A و B نشان می‌دهد که جدول توافق دو صفت متغیّر A و B می‌باشد.

بیان هندسی صفات کیفی

برای نمایش هندسی یا نمودار آماری صفات کیفی، از نمودار دایره‌ای یا قطاعی استفاده می‌شود.

ــ نمودار دایره‌ای یا قطاعی، شکل دیگری از نمودارهای آماری است که برای نمایش سهم هریک از حالتهای صفت کیفی در جامعه، به‌کار می‌رود.

برای رسم نمودار دایره‌ای، یک دایره با شعاع دلخواه درنظر گرفته، فراوانی یا درصد فراوانی هریک از حالات صفت را برحسب درجه، مطابق فرمول زیر محاسبه می‌کنیم و آن‌را به‌صورت کمانی از دایره، رسم می‌نماییم :

$$\frac{F_i \times ۳۶۰}{N}$$

مثال۸: نمودار دایره‌ای را برای هزینه خانوارها که از هزینه‌های مختلف مانند : هزینه‌خوراک، هزینه پوشاک، هزینه مسکن، هزینه بهداشت و درمان، هزینه آموزش و هزینه‌های متفرقه، تشکیل شده است و ارقام آن در جدول زیر داده شده، رسم می‌کنیم : (ارقام هزینه، متوسط هزینه‌های خانوارهای یک منطقه می‌باشند).

هزینه‌های خانوار	هزینه‌خوراک	هزینه‌پوشاک	هزینه‌مسکن	هزینه‌بهداشت‌ودرمان	هزینه‌آموزش	هزینه‌های متفرقه
مبلغ برحسب هزارتومان	۲۸	۱۰	۳۰	۵	۷	۲۰

ابتدا، هزینه‌ها را برحسب درجه، محاسبه می‌کنیم. در اینجا $N=۱۰۰$ می‌باشد.

$$\frac{۲۸ \times ۳۶۰}{۱۰۰} = ۱۰۰/۸^\circ$$

$$\frac{۱۰ \times ۳۶۰}{۱۰۰} = ۳۶^\circ$$

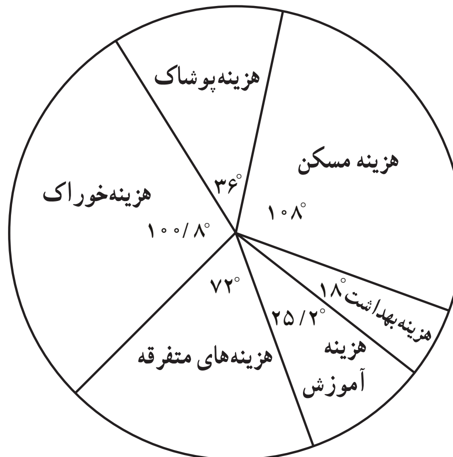
$$\frac{۳۰ \times ۳۶۰}{۱۰۰} = ۱۰۸^\circ$$

$$\frac{۵ \times ۳۶۰}{۱۰۰} = ۱۸^\circ$$

$$\frac{۷ \times ۳۶۰}{۱۰۰} = ۲۵/۲^\circ$$

$$\frac{۲۰ \times ۳۶۰}{۱۰۰} = ۷۲^\circ$$

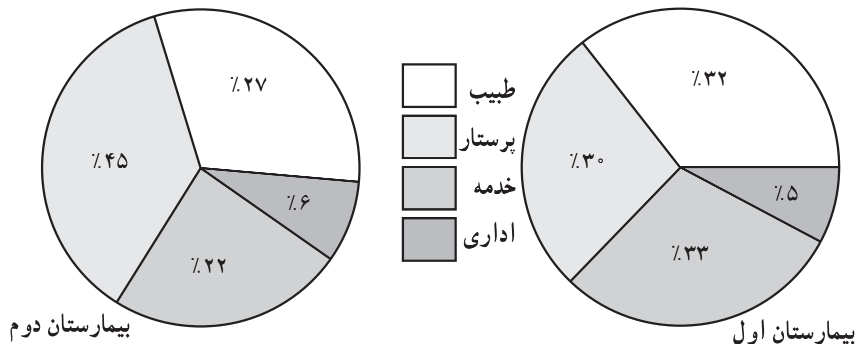
باید توجه نمود که جمع درجه‌ها بایستی °۳۶۰ باشد.



شکل ۱ــ نمودار دایره‌ای هزینه خانوارها

مناسب است، بخشهای دایره را با رنگهای مختلف و یا با هاشورهای متفاوت ازهم تفکیک کرد.

از نمودار دایره‌ای برای مقایسه ساختار دو جامعه (یا بیشتر) نیز استفاده می‌شود. به‌عنوان مثال، ترکیب خدمه دو بیمارستان را به‌منظور مقایسه، به‌صورت نمودار قطاعی می‌آوریم (شکل ۵ــ۲) :



از نمودار قطاعی برای بیان هندسی صفات متغیّر کمّی نیز استفاده می‌شود.

همین‌طور می‌توان از نمودارهای دیگری مانند نمودار ستونی برای بیان هندسی صفت متغیّر کیفی نیز استفاده نمود.

سؤالها و تمرینها

۱ــ نتایج مشاهدات روی صفات کیفی، چگونه بیان می‌شود؟

۲ــ برای گروه بندی نتایج مشاهدات صفات کیفی، چگونه عمل می کنند؟

۳ــ اگر صفت کیفی در جامعه، حالتهای مختلفی را اختیار کند، جدول توزیع صفت متغیّر، به چه صورت بیان می‌شود؟

۴ــ برای تبدیل صفات کیفی به صفات کمّی، از چه شیوه‌ای استفاده می‌شود؟

۵ــ مشخص‌کننده‌های عددی صفت کیفی، چگونه محاسبه می‌شوند؟

۶ــ میانگین صفت کیفی برابر با چه مقداری است؟

۷ــ واریانس صفت کیفی از چه رابطه‌ای محاسبه می‌شود؟

۸ــ کمترین مقدار واریانس صفت کیفی، چیست؟

۹ــ بیشترین مقدار واریانس صفت کیفی، چیست؟

۱۰ــ همگنی کامل، چه زمانی روی می‌دهد؟

۱۱ــ ناهمگنی کامل، چه زمانی روی می‌دهد؟

۱۲ــ برای مقایسه جامعه‌ها ازنظر صفات کیفی، از چه مفهومی استفاده می کنند؟

۱۳ــ توزیع گروه خونی جمعیّت دو منطقه، به‌صورت زیر به‌دست آمده است :

منطقه II

گروه خونی	**A**	**B**	**AB**	**O**
فراوانی‌نسبی	۰/۱۸	۰/۲۲	۰/۳۰	۰/۳۰

منطقه I

گروه خونی	**A**	**B**	**AB**	**O**
فراوانی‌نسبی	۰/۲۰	۰/۱۵	۰/۲۵	۰/۴۰

کدام‌یک از دو منطقهٔ مذکور، ازنظر گروه خونی O، همگن‌ترند؟

۱۴ــ اگر دو صفت کیفی با حالات متضاد، در یک جامعه برای عناصر آن مشاهده شده باشد، نتایج آن مشاهدات، چگونه بیان می‌شود؟

۱۵ــ این گروه‌بندی را چه می‌نامند؟

۱۶ــ جدول چهارخانه‌ای برای چه‌منظوری تشکیل می‌شود؟

۱۷ــ اگر دو صفت کیفی A و B، که هریک از آنها برای عناصر جامعه با بیش از دو حالت مشاهده شوند، نتیجه چنین مشاهداتی، چگونه گروه‌بندی می گردند؟

۱۸ـ در یک استان، ۴ نوع نوشابه تولید می‌شود، که با علامتهای A_1 ، A_2 ، A_3 و A_4 مشخص‌شده‌اند. از ۴۰ دانش‌آموز پسر و ۴۰ دانش‌آموز دختر نظرخواهی شد که چه نوع نوشابه‌ای را می‌پسندند، نتایج به‌قرار زیر به‌دست آمده است :

دختران	پسران
$A_1 = ۱۰$	$A_1 = ۱۲$
$A_2 = ۱۴$	$A_2 = ۱۰$
$A_3 = ۸$	$A_3 = ۱۵$
$A_4 = ۸$	$A_4 = ۳$

نتایج را در یک جدول دو بُعدی، گروه‌بندی نموده، جدول فراوانیهای نسبی را تشکیل دهید.

۱۹ـ درصد انواع محصولات کشاورزی تولید شده در یک استان به‌قرار زیر است :

نوع محصول	گندم	جو	ذرت	حبوبات
درصد	۴۵	۲۲	۱۷	۱۶

برای بیان ترکیب محصول تولید شده در این استان، نمودار قطاعی رسم کنید.

۲۰ـ درصد وامهای پرداختی بانک A در سالهای ۱۳۷۳ و ۱۳۷۵ به‌صورت زیر بوده است (ارقام ساختگی است) :

نوع وام	سال ۱۳۷۳	سال ۱۳۷۵
ساختمان	۳۱	۲۴
صنعت	۳۰	۲۸
کشاورزی	۲۰	۲۸
خدمات	۱۹	۲۰

برای مقایسه وامهای پرداختی در سالهای ۱۳۷۳ و ۱۳۷۵ نمودار دایره‌ای رسم کنید.

فصل ششم

توزیع مشترک دو صفت متغیر

هدفهای رفتاری: در پایان این فصل، فراگیر باید بتواند :

۱ـ جامعه را برحسب دو صفت متغیر کمی گروه‌بندی نماید.

۲ـ جدول توزیع فراوانی‌های دو صفت متغیر را تشکیل دهد.

۳ـ توزیعهای حاشیه‌ای x و y را به‌دست آورد.

۴ـ توزیعهای شرطی y برحسب x و x برحسب y را به‌دست آورد.

۵ ـ مشخصه‌های عددی برای دو صفت متغیر را به تفکیک محاسبه نماید.

۶ ـ مفهوم پراکندگی توأم دو صفت متغیر را بیان نماید.

در فصلهای گذشته، راجع به مطالعه جامعه‌ها با یک صفت متغیر بحث کردیم. ولی اغلب برای یک محقق نیاز می‌شود که اعضاء جامعه را توسط چند صفت متغیر با هم مورد مطالعه قرار دهد. در این فصل فقط به مطالعه دو صفت متغیر از اعضاء جامعه می‌پردازیم.

توزیع مشترک دو صفت متغیر x و y

اگر برای اعضاء جامعه دو صفت متغیر x و y با هم اندازه‌گیری شوند، آنگاه برای هریک از اعضاء یک زوج اندازه یا مقدار به‌دست خواهد آمد. مجموع زوجهای به‌دست آمده از نتایج مشاهدات مقادیر دو صفت را می‌توان به‌صورت زیر نشان داد.

شماره اعضاء	۱	۲	۳	n
اندازه صفت x	$x_۱$	$x_۲$	$x_۳$	x_n
اندازه صفت y	$y_۱$	$y_۲$	$y_۳$	y_n

چنانچه حجم جامعه بزرگ باشد، به‌طوری که هر زوج از اندازه‌های x و y بیش از یک بار

تکرار شوند، لزوماً نتایج مشاهدات را در جدول توزیع فراوانی دو بعدی وارد می‌کنند که در آن فراوانیها (F) با دو اندیس i برای صفت x و j برای صفت y مشخص می‌گردند، یعنی به‌صورت F_{ij} نشان داده می‌شوند. مجموعه دو صفت متغیر را دستگاه دو متغیر نیز می‌نامند. جدول ۱ توزیع فراوانیها را برای دو صفت متغیر x و y نشان می‌دهد :

جدول ۱

y \ x	y_1	y_2	......	y_j	......	y_t	$F_{i\circ}$
x_1	F_{11}	F_{12}	......	F_{1j}	......	F_{1t}	$F_{1\circ}$
x_2	F_{21}	F_{22}	......	F_{2j}	......	F_{2t}	$F_{2\circ}$
x_i	F_{i1}	F_{i2}	......	F_{ij}	......	F_{it}	$F_{i\circ}$
x_s	F_{s1}	F_{s2}	......	F_{sj}	......	F_{st}	$F_{s\circ}$
$F_{\circ j}$	$F_{\circ 1}$	$F_{\circ 2}$	......	$F_{\circ j}$	......	$F_{\circ t}$	n

در این جدول اگر فراوانیهای درون جدول (F_{ij}) را روی سطر جمع کنیم خواهیم داشت :

$$\sum_{j=1}^{t} F_{ij} = F_{i1} + F_{i2} + \dots + F_{it} = F_{i\circ}$$

چون فراوانیها روی اندیس j جمع شده‌اند بنابراین جواب حاصل به‌صورت $F_{i\circ}$ نوشته خواهد شد. هم‌چنین اگر فراوانیها را روی اندیس i جمع ببندیم جواب حاصل به‌صورت $F_{\circ j}$ نوشته خواهد شد.

$$\sum_{i=1}^{s} F_{ij} = F_{1j} + F_{2j} + \dots + F_{sj} = F_{\circ j}$$

بنابراین جمع سطرهای جدول به‌صورت $F_{1\circ}$ ، $F_{2\circ}$ ، ....، $F_{s\circ}$ درخواهد آمد و اگر آنها را روی اندیس i جمع کنیم حاصل حجم جامعه یعنی n را خواهد داد. و اگر $F_{\circ j}$ ها را روی j جمع کنیم نتیجه برابر حجم جامعه یعنی n خواهد شد.

$$\sum_{i=1}^{s} F_{i\circ} = F_{1\circ} + F_{2\circ} + \dots + F_{s\circ} = n$$

$$\sum_{j=1}^{t} F_{\circ j} = F_{\circ 1} + F_{\circ 2} + \dots + F_{\circ t} = n$$

مثال ۱: فرض کنید از جامعه دانش‌آموزان یک دبیرستان دو صفت متغیر x و y، که در آن x سن دانش‌آموزان و y تعداد افراد خانوار دانش‌آموزان، پرسش گردیده نتایج به‌صورت جدول توزیع فراوانیها (جدول۲) به‌دست آمده باشد.

جدول ۲

y تعداد افراد خانوار / x سن	۲	۳	۴	۵	۶	جمع سطر
۱۴	۳	۴	۶	۱۲	۵	۳۰
۱۵	۲	۷	۱۰	۱۱	۵	۳۵
۱۶	۴	۵	۹	۱۰	۴	۳۲
۱۷	۶	۴	۱۲	۱۰	۵	۳۷
۱۸	۳	۳	۷	۸	۷	۲۸
۱۹	۲	۶	۵	۴	۱	۱۸
جمع ستون	۲۰	۲۹	۴۹	۵۵	۲۷	۱۸۰

اعداد داخل جدول نشانه فراوانیهای هر زوج از دو صفت متغیر می‌باشد. برای مثال ۴ دانش‌آموز ۱۴ ساله تعداد نفرات خانوارشان ۳ نفره است و هم‌چنین ۱۰ دانش‌آموز ۱۵ ساله تعداد نفرات خانوارشان ۴ نفره می‌باشد.

توزیعهای حاشیه‌ای

از این جدول دوبعدی می‌توان توزیع هریک از صفات متغیر را جداگانه به‌دست آورد. مثلاً بدون درنظر گرفتن صفت y، می‌توان توزیع صفت متغیر x را به‌دست آورد.

جدول ۳

x	x_1	x_2	......	x_i	......	x_s	جمع
فراوانی $F_{i\circ}$	$F_{1\circ}$	$F_{2\circ}$	......	$F_{i\circ}$	......	$F_{s\circ}$	n

برای مثال ۱ توزیع سن دانش‌آموزان را بدون درنظر گرفتن تعداد افراد خانوار تشکیل می‌دهیم :

جدول ۴

سن x	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	جمع
فراوانی $F_{i\circ}$	۳۰	۳۵	۳۲	۳۷	۲۸	۱۸	۱۸۰

چون فراوانیها برای مقادیر صفت x (سن) از ستون کناری جدول قرار داده شده، به چنین توزیعی، توزیع کناری یا توزیع حاشیه‌ای x گویند. این توزیع نشان می‌دهد که توزیع سن دانش‌آموزان این دبیرستان چگونه است. همین‌طور می‌توان توزیع صفت y را به‌طور جداگانه از روی جدول دوبعدی ۱ به‌دست آورد :

جدول ۵

y	y_1	y_2	......	y_j	......	y_t	جمع
فراوانی $F_{\circ j}$	$F_{\circ 1}$	$F_{\circ 2}$	......	$F_{\circ j}$	......	$F_{\circ t}$	n

چنین توزیعی را توزیع حاشیه‌ای y نامند. چون در برابر مقادیر صفت y فراوانیها را از سطر پایین جدول ۱ قرار داده‌ایم.

توزیع حاشیه‌ای y را برای مثال ۱ به‌صورت جدول (۶) زیر می‌نویسیم :

جدول ۶

y (تعداد افراد خانوار)	۲	۳	۴	۵	۶	جمع
فراوانی $F_{\circ j}$	۲۰	۲۹	۴۹	۵۵	۲۷	۱۸۰

این توزیع به ما نشان می‌دهد که دانش‌آموزان این دبیرستان از نظر تعداد افراد خانوارشان چگونه توزیع شده‌اند.

توزیعهای شرطی

هرگاه بخواهیم برای اندازه معینی از صفت y، توزیع صفت x را بدانیم و یا برعکس برای اندازه یک x معین، توزیع صفت y را به‌دست آوریم. می‌توان با قرار دادن مقادیر صفت x در جدول و فراوانیهای متناظر با y معین را در برابر آنها، توزیع صفت x برحسب y معین را به‌دست آورد و آن‌را به‌صورت $X|Y\,.\,y_j$ نوشت و برای توزیع y برحسب x معین آن‌را به‌صورت $Y|X\,.\,x_i$ می‌نویسیم. به چنین توزیعهایی، توزیع شرطی x برحسب y و توزیع شرطی y برحسب x گویند. مثلاً اگر بخواهیم توزیع سن دانش‌آموزان را برای خانوارهای ۴ نفره به‌دست آوریم (با توجه به جدول ۲) خواهیم داشت :

جدول ۷

جمع	۱۹	۱۸	۱۷	۱۶	۱۵	۱۴	$X\|y\,.\,۴$
۴۹	۵	۷	۱۲	۹	۱۰	۶	$F_{i۴}$ فراوانی

ملاحظه می‌شود که فراوانیها از ستون $y\,.\,۴$ در برابر مقادیر صفت x نوشته شده است.

هم‌چنین اگر توزیع نفرات خانوار برای دانش‌آموزان ۱۷ ساله موردنظر باشد، می‌توان آن‌را به‌صورت زیر به‌دست آورد :

جدول ۸

جمع	۶	۵	۴	۳	۲	$Y\|x\,.\,۱۷$
۲۷	۵	۱۰	۱۲	۴	۶	$F_{۱۷j}$ فراوانی

در اینجا نیز فراوانیها از سطر $x\,.\,۱۷$ در برابر مقادیر y نوشته شده است.

از توزیعهای شرطی برای پی بردن به وجود بستگی بین دو صفت متغیر x و y استفاده می‌شود.

مشخصه‌های عددی برای توزیع مشترک دو صفت متغیر x و y

برای محاسبه میانگین صفت متغیر x و میانگین صفت متغیر y از توزیعهای حاشیه‌ای x و y استفاده نموده و مطابق طریقه محاسبه میانگین حسابی، می‌توان آنها را به‌دست آورد :

فرمولهای محاسباتی میانگینها به قرار زیر است :

$$\overline{X} = \frac{\sum_{i=۱}^{s} F_{i\circ} X_i}{n} \qquad (۱)$$

$$\overline{Y} = \frac{\sum_{j=۱}^{t} F_{\circ j} Y_j}{n} \qquad (۲)$$

برای مثال ۱، میانگین صفت متغیر x (سن دانش‌آموزان) و میانگین صفت متغیر y (تعداد افراد خانوار) را محاسبه می‌کنیم :

$$\overline{X} = \frac{\sum_{i=۱}^{s} F_{i\circ} X_i}{n} = \frac{۳۰ \times ۱۴ + ۳۵ \times ۱۵ + ۳۲ \times ۱۶ + ۳۷ \times ۱۷ + ۲۸ \times ۱۸ + ۱۸ \times ۱۹}{۱۸۰}$$

$$= \frac{۲۹۳۲}{۱۸۰} = ۱۶/۲۹$$

به عبارت بهتر میانگین سن دانش‌آموزان ۱۶/۲۹ سال می‌باشد.

$$\overline{Y} = \frac{\sum_{j=۱}^{t} F_{\circ j} Y_j}{n} = \frac{۲۰ \times ۲ + ۲۹ \times ۳ + ۴۹ \times ۴ + ۵۵ \times ۵ + ۲۷ \times ۶}{۱۸۰}$$

$$= \frac{۷۶۰}{۱۸۰} = ۴/۲۲$$

یعنی متوسط تعداد افراد خانوار دانش‌آموزان ۴/۲۲ نفر می‌باشد.

هم‌چنین می‌توان پراکندگی (واریانس) هریک از صفات متغیر x و y را توسط روابط زیر محاسبه نمود :

$$V(X) = \frac{\sum_{i=۱}^{s} F_{i\circ} X_i^۲}{n} - \overline{X}^۲ \qquad (۳)$$

$$V(Y) = \frac{\sum_{j=۱}^{t} F_{\circ j} Y_j^۲}{n} - \overline{Y}^۲ \qquad (۴)$$

برای مثال ۱ واریانس صفت متغیر X و واریانس صفت متغیر Y را محاسبه می‌کنیم :

$$V(X)=\frac{\sum_{i=۱}^{s}F_{i\circ}X_i^{۲}}{n}-\overline{X}^{۲}$$

$$=\frac{۳۰\times ۱۴^{۲}+۳۵\times ۱۵^{۲}+۳۲\times ۱۶^{۲}+۳۷\times ۱۷^{۲}+۲۸\times ۱۸^{۲}+۱۸\times ۱۹^{۲}}{۱۸۰}-(۱۶/۲۹)^{۲}$$

$$=\frac{۴۸۲۱۰}{۱۸۰}-۲۶۵/۳۶=۲۶۷/۸۳-۲۶۵/۳۶=۲/۴۷$$

$$V(Y)=\frac{\sum_{j=۱}^{t}F_{\circ j}Y_j^{۲}}{n}-\overline{Y}^{۲}$$

$$=\frac{۲۰\times ۲^{۲}+۲۹\times ۳^{۲}+۴۹\times ۴^{۲}+۵۵\times ۵^{۲}+۲۷\times ۶^{۲}}{۱۸۰}-(۴/۲۲)^{۲}$$

$$=\frac{۳۴۷۲}{۱۸۰}-۱۷/۸۱=۱۹/۲۹-۱۷/۸۱=۱/۴۸$$

علاوه بر پراکندگی صفات متغیر x و y، پراکندگی توأم دو صفت متغیر x و y را نیز محاسبه می‌کنند که آن‌را کواریانس x و y نامند و به‌صورت Cov(x,y) نوشته می‌شود. فرمول محاسبه کواریانس x و y به قرار زیر است :

$$Cov(x,y)=\frac{\sum_i\sum_j F_{ij}X_iY_j}{n}-\overline{X}.\overline{Y} \qquad (۵)$$

از کواریانس x و y برای آگاهی از همبستگی بین دو صفت x و y استفاده می‌کنند.

هرگاه مقدار کواریانس x و y برابر صفر شود یعنی بین دو صفت x و y همبستگی خطی وجود ندارد و هرگاه مقدار کواریانس x و y بزرگتر از صفر (مثبت) باشد یعنی بین دو صفت x و y همبستگی به‌طور مستقیم وجود دارد. و چنانچه کوچکتر از صفر (منفی) باشد یعنی بین دو صفت x و y همبستگی به‌طور معکوس وجود دارد.

برای مثال ۱ کواریانس x و y را محاسبه می‌کنیم :

$$Cov(x,y)=\frac{\sum_i\sum_j F_{ij}X_iY_j}{n}-\overline{X}\overline{Y}$$

$$=\underline{۳\times ۱۴\times ۲+۴\times ۱۴\times ۳+۶\times ۱۴\times ۴+۱۲\times ۱۴\times ۵+۵\times ۱۴\times ۶+۲\times ۱۵\times ۲+}$$

۷. ۱۵. ۳. ۱۰. ۱۵. ۴. ۱۱. ۱۵. ۵. ۵. ۱۱. ۶. ۴. ۱۶. ۲. ۵. ۱۶. ۳.

۹. ۱۶. ۴. ۱۰. ۱۶. ۵. ۴. ۱۶. ۶. ۶. ۱۷. ۲. ۴. ۱۷. ۳. ۱۲. ۱۷. ۴.

۱۰. ۱۷. ۵. ۵. ۱۷. ۶. ۳. ۱۸. ۲. ۳. ۱۸. ۳. ۷. ۱۸. ۴. ۸. ۱۸. ۵

. ۷. ۱۸. ۶. ۲. ۱۹. ۲. ۶. ۱۹. ۳. ۵. ۱۹. ۴. ۴. ۱۹. ۵. ۱. ۱۹. ۶

۱۸۰

. ۱۲۳۵۲/۱۸۰ . ۱۶/۲۹. ۴/۲۲. ۶۸/۶۲. ۶۸/۷۴.. ۰/۱۲

برای سهولت محاسبات می‌توان کلیه محاسبات را توسط جدول زیر انجام داد : (جدول ۹)

جدول ۹

x \ y	۲	۳	۴	۵	۶	$F_{i\circ}$	$F_{i\circ}X_i$	$F_{i\circ}X_i^2$	$_j F_{ij}Y_j$	$X_i . F_{ij}Y_j$
۱۴	۳	۴	۶	۱۲	۵	۳۰	۴۲۰	۵۸۸۰	۱۳۲	۱۸۴۸
۱۵	۲	۷	۱۰	۱۱	۵	۳۵	۵۲۵	۷۸۷۵	۱۵۰	۲۲۵۰
۱۶	۴	۵	۹	۱۰	۴	۳۲	۵۱۲	۸۱۹۲	۱۳۳	۲۱۲۸
۱۷	۶	۴	۱۲	۱۰	۵	۳۷	۶۲۹	۱۰۶۹۳	۱۵۲	۲۵۸۴
۱۸	۳	۳	۷	۸	۷	۲۸	۵۰۴	۹۰۷۲	۱۲۵	۲۲۵۰
۱۹	۲	۶	۵	۴	۱	۱۸	۳۴۲	۶۴۹۸	۶۸	۱۲۹۲
$F_{\circ j}$	۲۰	۲۹	۴۹	۵۵	۲۷	۱۸۰	۲۹۳۲	۴۸۲۱۰	۷۶۰	۱۲۳۵۲
$F_{\circ j}y_j$	۴۰	۸۷	۱۹۶	۲۷۵	۱۶۲	۷۶۰				
$F_{\circ j}y_j^2$	۸۰	۲۶۱	۷۸۴	۱۳۷۵	۹۷۲	۳۴۷۲				

محاسبه اندازه همبستگی بین دو صفت متغیر x و y و تفسیر آن‌را در جلد دوم کتاب روشهای آماری از نظر خواهید گذراند.

سؤالها و تمرینها

۱ـ توزیع مشترک دو صفت متغیر چیست؟

۲ـ هدف از مطالعه جامعه‌ها با بیش از یک صفت متغیر چیست؟

۳ـ فراوانی گروه ijام چیست؟

۴ـ توزیع حاشیه‌ای صفت متغیر x چگونه به‌دست می‌آید؟

۵ـ توزیع حاشیه‌ای صفت متغیر y چگونه به‌دست می‌آید؟

۶ـ توزیع شرطی صفت متغیر x برحسب y چه مفهومی دارد؟

۷ـ توزیع شرطی صفت متغیر y برحسب x چه مفهومی دارد؟

۸ـ از توزیع‌های شرطی برای چه منظوری استفاده می‌شود؟

۹ـ در توزیع مشترک دو صفت متغیر، میانگین حسابی دو صفت چگونه محاسبه می‌شود؟

۱۰ـ در توزیع مشترک دو صفت متغیر، پراکندگی صفات چگونه محاسبه می‌شود؟

۱۱ـ کواریانس دو صفت متغیر چیست؟ فرمول آن‌را بنویسید.

۱۲ـ از کواریانس برای آگاهی از چه خصوصیاتی از دو صفت متغیر استفاده می‌کنند؟

۱۳ـ هرگاه کواریانس دو صفت متغیر برابر صفر گردد چه مفهومی دارد؟

۱۴ـ نتایج مشاهدات دو صفت x و y در جدول زیر آمده است :

x	۸	۱۰	۱۱	۱۲	۱۴
y	۱۰	۶	۷	۳	۵

میانگین x و میانگین y را محاسبه کنید.

واریانس x و واریانس y را محاسبه کنید.

کواریانس x و y را محاسبه کنید.

۱۵ــ توزیع مشترک دو صفت متغیر x و y به‌شرح جدول زیر است :

x \ y	۳	۴	۵	۶	$F_{i\circ}$
۵	۲	۳	۳	۲	۱۰
۱۰	۳	۱۰	۱۰	۲	۲۵
۱۵	۱	۲	۷	۵	۱۵
$F_{\circ j}$	۶	۱۵	۲۰	۹	۵۰

مطلوبست :

الف ــ توزیع حاشیه‌ای x و توزیع حاشیه‌ای y

ب ــ توزیع شرطی y برحسب ۱۰ . x

ج ــ توزیع شرطی x برحسب ۵ . y

د ــ میانگین x و میانگین y

هـ ــ واریانس x و واریانس y

و ــ کواریانس x و y و تفسیر آن

منابع و مآخذ

۱ــ مفاهیم اساسی آمار تألیف علی مدنی ــ انتشارات فروردین.

۲ــ آنالیز آماری (جلد اول) تألیف علی مدنی.

۳ــ آمار توصیفی تألیف ژرار کالت ــ ترجمه حسن صادقی از انتشارات دانشگاه فردوسی مشهد.

۴ــ آمار و کاربرد آن در مدیریت تألیف مهدی صفاری ــ نشر دانا.

۵ــ آمار مقدماتی (جلد اول) تألیف وونا کات ــ ترجمه دکتر مشکانی.

۶ــ Statistical Methods Snedecor & Cochran.

۷ــ Statistical Analysis for Managerial Decisions Boot and Cox.

۸ــ Introduction to Mathematical Statistics Paul . G . Hoel.

